

**Universidad de Chile**

Facultad de Filosofía y Humanidades

Escuela de Postgrado

**Hacia una Teoría de la Naturalización de la  
Intencionalidad**

(Tesis para optar al grado de Magíster en Filosofía con Mención en  
Epistemología)

**Alumno:**

Matías Bulnes

**Profesores Guía:**

Alejandro Ramírez, Guido Vallejos

**Santiago, 2004**

**Prólogo**

La perspicacia de Quevedo tan sólo admitió la sospecha de que nada se sabe. Afirmar, en cambio, esto último hubiese significado su inmediata invalidación. Quine y Laudan llegaron a una conclusión similar por caminos muy distintos: que no hay verdades absolutas sino que lo que se sabe, o se cree saber, depende de las restantes cosas que se saben o cree saber. El conocimiento es, pues, como una gran red de enunciados cada uno entrelazado con los demás de modo que no se puede jalar de uno sin mover todo el conjunto. El peso del conjunto es lo que lo que sostiene a sus partes. Esto es lo que se conoce comúnmente por Holismo de la Confirmación. El vasto horizonte de sus implicancias filosóficas aun no puede vislumbrarse.

En lo que a mí respecta, lo verdaderamente interesante de la tesis anterior no está tanto en su aplicación a la ciencia como en su aplicación al conocimiento en general, del cual la ciencia es parte significativa, pero parte al fin. Su utilidad surgió de improviso en una conversación con unos amigos científicos. Uno de ellos sugirió que el problema de la filosofía es que, al no

tener un soporte empírico, nunca se puede saber si una teoría vale o no la pena. Así, la filosofía y, en general, todo pensamiento alejado de la posibilidad de contrastación empírica pareciera ser una especie de pseudo-conocimiento, cuya dignidad no supera la de charlatanes y timadores. Esto, sin embargo, contrastaba con la impresión que mi aun incipiente experiencia en filosofía analítica y epistemología habían dejado. Entonces me dispuse a dilucidar cómo una disciplina podía poseer criterios evaluativos de teorías independientes que la hagan una disciplina valiosa sin rozar en ningún momento la contrastación empírica. La respuesta, repito, la encontré en el Holismo de la Confirmación.

La versión de Laudan de esta tesis epistemológica es que son al menos dos los factores evaluativos que determinan si una teoría será aceptada al interior de una comunidad. Éstos son su contrastabilidad empírica y su relación con otras teorías circundantes, es decir, su compatibilidad con otros postulados teóricos anexos. Así, este último ítem resultaría un bien de alto valor gnoseológico, al punto que sería, en principio al menos, admisible que la comunidad aceptara una teoría sólo por su aporte en la coordinación de otras teorías aun cuando no contara con ningún grado de contrastabilidad empírica. Llevando esta idea a la metáfora de la red de enunciados antes mencionada, podríamos imaginar que los cabos de la red que quedan sueltos son aquellos donde la contrastación tiene lugar y que en último término sostienen toda la estructura. Sin embargo, es perfectamente posible que existan porciones interiores de la red, esto es, sin ningún cabo suelto, que, a pesar de no ofrecer soporte empírico, pueden ser fundamentales al momento de mantener la estructura en pie.

La filosofía es, según creo, una disciplina de este tipo. El rol de las teorías filosóficas es hacer que otras disciplinas que sí ofrecen contrastación empírica queden bien unidas en sus bordes. Las teorías filosóficas son porciones unitarias interiores a la gran red que es el conocimiento. Entre los nodos que las componen se encuentran los presupuestos ontológicos fundamentales los cuales, a su vez, ofrecen una pauta flexible de ajuste para otras teorías periféricas. Acaso tienen razón los físicos cuánticos que reclaman que no es tarea de ellos compatibilizar la Mecánica Cuántica con los presupuestos ontológicos fundamentales del determinismo causal. Acaso corresponda a los filósofos cambiar dichos presupuestos de modo que la Mecánica Cuántica pegue bien con las restantes teorías científicas. Sea cual sea la verdad, hasta el momento no se ha tenido éxito.

Como sea, el tema de la presente investigación es el de la intencionalidad y el problema Mente/Cuerpo. Esto pareciera no tener ninguna relación con lo que hemos dicho hasta aquí.

Sin embargo, pocas dudas pueden haber respecto de que el problema Mente/Cuerpo ha persistido por siglos como un desbarajuste en la gran red del conocimiento. Intentar atacar dicho problema, ya sea por vías de la intencionalidad o por cualquiera otra, supone unir una serie de cabos sueltos al interior de nuestra red de conocimiento contemporánea. Naturalmente, no espere el lector encontrarse con abundante contrastación empírica a lo largo de la presente tesis, sino que considérese satisfecho si lo que lee le parece compatible con el entorno teórico en que pretendemos situarnos, a saber, el de las ciencias naturales, el materialismo ontológico y la filosofía analítica en general. Si logro causar esta impresión en mis lectores habré alcanzado mi objetivo.

Matías Bulnes B.

# Capítulo 1

Veinticinco siglos de filosofía no han bastado para doblegar el problema Mente/Cuerpo<sup>1</sup>. Durante este largo periodo del hombre en la tierra el problema ha rondado constantemente en el pensamiento de muchas generaciones de filósofos. Cada una de éstas ha esbozado distintas formulaciones del mismo problema, sin perjuicio de que todas ellas compartan una pregunta fundamental que permite que muy diversos planteamientos sean reunidos bajo un mismo nombre. Sin lugar a dudas, esta pregunta ha de ubicarse entre las grandes de la filosofía universal y, probablemente, como suele ocurrir en estos casos, las respuestas que puedan entregarse no constituyan certezas absolutas sino que estén sujetas a los presupuestos fundamentales propios de cada momento histórico.

En una formulación suficientemente general que permita incluir posturas escépticas dentro de las respuestas posibles, la pregunta debe tener un carácter personal. Ésta sería ¿De cuántos dominios independientes estoy hecho? O bien, ¿cuántas sustancias ontológicas independientes se requieren para explicar toda mi fenomenología<sup>2</sup>? Aquí la noción de independencia entre dominios o sustancias ontológicas no supone ausencia de interacción entre ellas, sino que, más bien, exige simplemente que uno no pueda construirse a partir del otro. ¿Quién se atrevería a negar que, supuesta la existencia de lo mental y lo material como dominios independientes, el primero no influye de algún modo en el segundo? Sin embargo, una respuesta dualista como ésta exige que ni lo mental se construya a partir de lo material ni viceversa.

---

<sup>1</sup> A pesar de que el problema Mente/Cuerpo como tal existe desde hace bastante menos que 25 siglos, uno podría decir que la problemática que éste plantea ha estado presente desde los albores de la filosofía. Lo medular del problema es la pregunta ontológica respecto de realidades independientes motivada por la percepción del lo mental. En adelante, cuando hablemos del problema Mente/Cuerpo nos estaremos refiriendo a esta problemática general.

<sup>2</sup> No es mi intención aludir al término “fenomenología” en el sentido de la tradición filosófica descendiente de Husserl. Por el contrario, ésta es la mejor expresión de que dispongo para referirme al conjunto de fenómenos asociados al acto consciente de la percepción.

En segundo lugar, qué soy yo o qué constituye mi fenomenología debe entenderse en la pregunta anterior como el conjunto de fenómenos puros que recibo por medio de mi percepción. En otras palabras, éstos son los hechos que percibo independientemente de mis compromisos teóricos con ciertas formas de interpretar dichos hechos. Podríamos considerar como un ejemplo cercano, experimentado por todo ser humano, aquello que percibe un niño suficientemente pequeño que no ha pasado por procesos de aprendizaje de teorías *explicativas* del mundo, culturalmente parciales. En un caso así, todo lo que llega al niño a través de los sentidos viene determinado por la estructura cognitivo-perceptual innata de los seres humanos y nada más. No hay, pues, fenómenos interpretados tras el prisma de una teoría, sino tan sólo fenómenos.

Una vez hechas estas aclaraciones, podemos encontrar distintas respuestas lógicamente factibles a tan ambiciosa pregunta, cada una de las cuales supone compromisos con distintos postulados ontológicos fundamentales. Así, una sería que estoy compuesto de dos sustancias independientes, una material y otra mental, de modo que fenómenos situados en estos dominios interaccionan para dar lugar a lo que soy. Esta respuesta supone, pues, inclinarse por una concepción ontológica (cuando menos) dualista. Entre sus defensores encontramos al célebre filósofo francés René Descartes. Pero incluso podríamos clasificar como adeptos a esta concepción a filósofos anteriores que no tocaron explícitamente el problema Mente/Cuerpo. Platón, por ejemplo, postulaba que el dominio de las ideas guardaba independencia de las vicisitudes del mundo exterior y, por lo tanto, suponía que las primeras no pueden construirse a partir de lo material, aun cuando tampoco negaba la existencia de esto último. En consecuencia, podría decirse que si Platón hubiese tratado explícitamente el problema probablemente se hubiese inclinado por una salida dualista.

Entre las respuestas monistas encontramos dos posibles. Una consiste en asumir una ontología materialista; esto es, plantear que todo aquello que nos constituye resulta de interacciones materiales o bien se deriva de ellas. Los llamados fenómenos mentales serían, pues, expresión de otros fenómenos materiales subyacentes y no siempre perceptibles a simple vista. Esta concepción es la que hoy por hoy se encuentra de moda, presumiblemente debido al grandioso éxito que dicho supuesto ontológico a traído a la ciencia moderna. La respuesta monista alternativa consistiría, en cambio, en asumir una ontología mentalista, según la cual sólo existen fenómenos mentales mientras que los materiales serían una expresión ilusoria de aquellos. Entre los defensores de esta concepción encontramos a filósofos como Berkeley,

partidarios de las llamadas corrientes escéptico-solipsistas, las cuales dudan de la existencia de un mundo exterior a la vida mental.

La dificultad que supone el problema así planteado es que el hombre está separado de la supuesta realidad exterior por un pozo de error. En otras palabras, no tenemos acceso directo a la realidad exterior sino que cualquier postulación de ella será indefectiblemente hipotética; y cualquier hipótesis, surgida del uso de nuestras facultades epistémicas, supone la posibilidad de error. Cada individuo cuenta, en cambio, con acceso directo (i.e. sin error) a los fenómenos mentales, sin mediar evidencia epistémica. Consecuentemente, los partidarios de posiciones solipsistas argumentan que aun la percepción sensorial del supuesto mundo material es un fenómeno mental y la existencia de este último tipo de fenómenos es de lo único de que podemos estar seguros.

Hume observó que este tipo de argumentos no admite la menor réplica y no produce la menor convicción. Desde la perspectiva del problema Mente/Cuerpo, sin embargo, la primera sentencia podría ser imprecisa. De hecho, existen réplicas posibles en contra de éstos. Consideremos, por ejemplo, aquella conocida como la de la “mayor utilidad”. Su idea es que, exista o no una realidad exterior al dominio mental, resulta más útil asumir que sí existe. Si no, intente el lector suponer lo contrario y verá como rápidamente vuelve usted a actuar, creencia o no mediante, como si hubiera un mundo exterior.

En segundo lugar, no es difícil percatarse de que el argumento que va desde la certeza del dominio mental a la inexistencia de la realidad (material) exterior es falaz. La razón es que el hecho de que no podamos tener certeza de la existencia del mundo exterior no excluye la posibilidad de que éste de hecho exista. Del mismo modo, la certeza de la existencia de lo mental no supone la certeza de la inexistencia de lo material. Consecuentemente, lo que sí se sigue de lo anterior es la incerteza de este tipo de argumentos solipsistas.

En el fondo de tan delicado asunto está el hecho de que si bien la postulación de la existencia de la realidad material está mediado por nuestras capacidades epistémicas (viz. teoría mediante) también lo está la postulación de su inexistencia. Por lo tanto, aun cuando podamos tener certeza de la existencia de fenómenos mentales debido a la experiencia directa, no-epistémica que tenemos de ellos; aun así, la existencia o inexistencia de la realidad material deberá decidirse inevitablemente en el dominio teórico. Del mismo modo, la independencia ontológica de lo mental como su posible reducción a lo material deberá también ser discutida

en ese ámbito. El problema Mente/Cuerpo sigue, pues, vigente a pesar de los argumentos escéptico-solipsistas.

## **Dualismo versus materialismo**

Más allá de cuán sólidos puedan sonar los argumentos que puedan esgrimirse en contra de las posturas escéptico-solipsistas, es un hecho que la tradición contemporánea de la filosofía ha considerado la realidad material como existente. Más aun, por lo general, en el ámbito de la filosofía analítica –como se ha llamado a aquellas más cercanas a la ciencia- se ha supuesto que la única realidad existente es la material. Esto no significa descartar la existencia de lo mental como fenómeno, sino que, más bien, supone que ello constituye una expresión de un sustrato material. En otras palabras, en la tradición analítica de la filosofía contemporánea es ampliamente aceptada la posibilidad, en principio, de reducir los fenómenos mentales a otros materiales de algún tipo<sup>3</sup>.

A pesar de esta abierta inclinación hacia la salida materialista del problema Mente/Cuerpo, la polémica suscitada entre sus defensores y los partidarios de la salida dualista resulta mucho más rica en varios aspectos que aquélla sostenida con los escéptico-solipsistas. Esto probablemente se deba a que esta discusión, aun cuando ontológica, está ciertamente asociada a la evidencia epistémica disponible, que es el modo como habitualmente hacemos teorías; mientras que la otra era una acerca de la validez de la evidencia epistémica versus otro tipo de evidencia acaecida en el contexto de la experiencia directa de nuestra actividad mental.

Como sea, uno podría preguntarse por qué la discusión respecto de qué estamos compuestos gira en torno a sólo dos sustancias posibles. En otras palabras, no deja de ser sugerente que parezca tan descabellado sostener, como hicieron algunos pensadores presocráticos, que no es una o dos las sustancias ontológicamente independientes de las cuales estamos compuestos, sino que tres, cuatro o diez. Y la razón puede encontrarse en el hecho de que, dado que la discusión está mediada por la evidencia epistémica disponible, ésta está sujeta a ciertas leyes

---

<sup>3</sup> No es mi intención entrar en la discusión respecto del eliminativismo. De hecho, postulo la aceptación de la reducción de lo mental a lo material sólo como una sentencia ontológica, aceptable por partidarios y detractores de posturas eliminativistas, pues no descarto *a priori* la posibilidad de que pueda excluirse lo mental en una explicación completa de la fenomenología humana.

heurísticas respecto del conocimiento. En este caso particular podríamos apelar a un principio de simpleza, habitualmente usado en la actividad cotidiana de científicos y pensadores, el cual postula que frente a dos teorías de un mismo fenómeno igualmente explicativas habrá de elegirse aquella que sea más simple. Consecuentemente, para el caso de la discusión del problema Mente/Cuerpo es claro que a mayor número de sustancias independientes que se postulen para explicar la fenomenología de nuestra composición humana mayor complejidad de la explicación de los fenómenos en cuestión. Y, por otro lado, es materia de consenso que con la postulación de dos basta. La pregunta que se abre es si bastará con la postulación de una.

Pues bien, la formulación del problema Mente/Cuerpo va tomando una nueva forma. La pregunta fundamental antes planteada respecto de qué estamos compuestos, puede reemplazarse en el presente contexto por cuántas sustancias ontológicamente independientes se requieren para explicar nuestra fenomenología. Por lo tanto, en virtud del principio heurístico de la simpleza, el problema de decidir entre una concepción materialista (viz. monista) y otra dualista puede reducirse al de la posibilidad de explicar los fenómenos mentales en términos materiales. Si es posible diseñar una teoría coherente de los primeros y de las leyes que los rigen como expresión de fenómenos y leyes estrictamente materiales (i.e. sin referencia a otros fenómenos mentales) entonces el problema Mente/Cuerpo se decidirá en favor del materialismo. Si no es posible realizar dicho ejercicio habrá que aceptar la solución dualista.

Acaso el mejor argumento con que cuentan los dualistas para postular la irreductibilidad de los fenómenos mentales a otros materiales sea la evidente diferencia entre las propiedades sensoriales de las entidades mentales y las de las entidades materiales. Así, por ejemplo, las últimas poseen una textura asimilable por medio del tacto, un olor aprehensible por medio del olfato, una forma geométrica reconocible por medio de la vista, etc. En oposición, las entidades mentales tales como los pensamientos o conceptos no poseen ninguna de dichas características. Y no es sólo que no podamos experimentar dichas propiedades –dirían los dualistas- sino que pareciera ser que simplemente no las poseen. Por otro lado, es un hecho difícilmente cuestionable que cualquier entidad que se construya a partir de otras materiales siempre poseerá propiedades sensoriales (comúnmente las mismas de sus constituyentes). Consecuentemente, si las entidades mentales no poseen propiedades sensoriales no podrían construirse a partir de otras que sí tienen, como serían cualquier clase de entidad material. Las entidades mentales no serían, pues, reducibles a otras materiales.

Sin embargo, este tipo de argumentos pasa por alto el mismo hecho que los escéptico-solipsistas usaban para debilitar a sus detractores; a saber, que los estados mentales y el aparato perceptivo son parte del mismo mecanismo cognitivo. De hecho, ésta es precisamente la razón por la cual nuestra experiencia de nuestros estados mentales no es mediada por evidencia epistémica alguna. Por el contrario, tenemos experiencia directa de nuestros estados mentales debido a que éstos son parte de lo que somos. De otro modo, si los estados mentales estuvieran escindidos de nuestra constitución individual indudablemente serían aprehensibles epistémicamente. Más aun, esto es justamente lo que sucede con los estados mentales ajenos. En la medida en que los estados mentales de mi amigo Juan no son parte de mi constitución personal sino que se encuentran fuera de mí, tengo sólo acceso epistémico (*a forteiori*, sujeto a error) a ellos. Cualquier cosa que pueda decir de ellos no dejará de pertenecer al dominio hipotético y será suscitada por mi percepción sensorial.

Conforme con esto, la intuición de que los estados mentales propios o ajenos no poseen las propiedades sensoriales habituales de las entidades materiales bien puede explicarse como un efecto ilusorio debido a una diferencia sustancial en el modo como experimentamos los primeros y los últimos. En efecto, los únicos estados mentales que experimentamos *qua* estados sin propiedades sensoriales son los propios. La atribución de estados similares a otros individuos es meramente hipotética, siendo el caso que si pudiésemos ver lo que ocurre dentro de la cabeza de otro individuo cuando le atribuimos dichos estados no veríamos más que sinápsis neuronales y descargas eléctricas, todas ellas con las propiedades sensoriales habituales en lo material. En cambio, el que los estados mentales propios parezcan carecer de las propiedades sensoriales de las entidades materiales no se debe a una diferencia ontológica entre los primeros y las últimas sino a una diferencia en el modo como los experimentamos. Mientras que las últimas son experimentadas por medio del aparato cognitivo (i.e. poseemos acceso epistémico) los primeros son experimentados directamente, sin mediación de éste.

Fodor observa que es una limitante de nuestra condición human no tener acceso metafísico a la realidad exterior, como tendría un supuesto dios panteísta. A esto habría que agregar que resulta nefasto el que no tengamos acceso epistémico a nuestra actividad mental pues esto sugiere la inconveniente ilusión de que ésta es de naturaleza distinta que la realidad exterior. En palabras simples, podríamos decir metafóricamente que del mismo modo como no podemos percibir un trozo de nuestra piel con el mismo trozo de nuestra piel, no podemos percibir nuestros estados mentales, parte fundamental de nuestro aparato cognitivo, con el mismo aparato cognitivo. Su función es permitir relacionarnos con el mundo exterior y,

obviamente, el aparato cognitivo no es exterior a sí mismo. En consecuencia, no estamos capacitados para relacionarnos con nuestro aparato cognitivo o partes de él del mismo modo como lo hacemos con los objetos del exterior.

Place (1956) toma una idea análoga para desarrollar un argumento en contra de ciertas posiciones cercanas al dualismo ontológico. Place llamaba a éstas consideraciones la Falacia Fenomenológica y argüía que, antes que a los filósofos, era a los fisiólogos a quienes ésta preocupaba mayormente (Sherrington 1947). Éstos postulaban que la imposibilidad del observador de entregar una descripción de sus estados mentales en términos de las propiedades fisiológicas del sistema nervioso central servía como prueba a favor de la tesis de que dichos estados son ontológicamente independientes del sistema nervioso central – representante del sustrato material. Así, por ejemplo, un sujeto describiendo la imagen que percibe del sol probablemente la referirá como un disco dorado ubicado en el cielo; y esto no puede ser derivado de los procesos fisiológicos presentes en el fenómeno de percepción, los que involucrarán una descripción físico-química del mismo (e.g. en términos de ondas electromagnéticas, reacciones de la pupila, etc.). En otras palabras, la descripción físico-química de los objetos y procesos de la percepción jamás podrá dar cuenta de los estados psicológicos, intencionales del observador asociados al mismo fenómeno perceptivo. Una cosa es que un objeto emita una cierta frecuencia de ondas electromagnéticas (descripción físico-química) y otra muy distinta es que el objeto sea azul (descripción intencional), dirían los fisiólogos antes aludidos.

Sin embargo, Place observa que tras este tipo de argumentos subyace el supuesto erróneo de que una descripción de un objeto corresponde a una descripción de nuestros estados psicológicos asociados a él. Para usar una metáfora bastante desacreditada en filosofía de la mente pero no por esto ineficaz en el presente contexto, si imaginamos que todo sujeto cognoscente tiene dentro de sí un pequeño hombre que ve en una pantalla en lenguaje intencional lo que nuestro aparato perceptivo capta del exterior en lenguaje físico-químico, lo anterior equivaldría a decir que una descripción de un objeto corresponde a una descripción de la imagen proyectada en la pantalla interior. Este supuesto dominio de fenómenos interiores para el cual no hay lugar en el mundo físico es lo que Place llama mundo fenomenológico, presumiblemente en alusión a ciertas doctrinas filosóficas de origen franco-alemán muy de moda durante buena parte del siglo XX. No obstante, no hay un mundo fenomenológico tal. Una descripción intencional de un objeto y otra físico-química son descripciones del mismo objetos y de las mismas propiedades metafísicas. No es que los

hombres vengamos de fábrica con la capacidad de mapear la realidad material en otra realidad fenomenológica (como sería la pantalla antes aludida) ontológicamente escindida de aquélla, sino que frente a la recepción de estímulos nerviosos asociados a ciertas ondas electromagnéticas provenientes de ciertos objetos aprendemos a referir esta propiedad del objeto como la de ser azul. La descripción intencional es una mera forma de referir la realidad material y no es síntoma de un dominio ontológico independiente de ésta.

En resumen, ninguno de los argumentos comúnmente esgrimidos para defender la existencia del dominio mental como ontológicamente independiente del material resulta convincente. Luego, en la medida en que el dilema central del problema Mente/Cuerpo había sido reducido al de cuántos dominios ontológicamente independientes se requieren para explicar los fenómenos humanos, podemos convenir en que aun no ha sido establecida la necesidad de apelar a dos dominios (el mental y el material) para resolverlo. Sin embargo, para nuestra desgracia, con la discusión anterior tampoco hemos demostrado la suficiencia del dominio material para dar cuenta de nuestra completa realidad humana. Para alcanzar este objetivo tendremos que ser capaces de explicar los fenómenos mentales en términos de una teoría materialista del mundo y del hombre. En lo que resta del presente capítulo daremos el primer paso en esta dirección al tratar de caracterizar los fenómenos mentales como fenómenos materiales con una propiedad adicional, a saber, la de poseer intencionalidad. Si tenemos éxito en esta tarea habremos reducido el problema al de explicar esta propiedad de poseer intencionalidad en términos materialistas. Éste será nuestro objetivo en los capítulos restantes de la presente investigación. Pero vamos paso a paso, pues quien mira demasiado lejos pierde de vista lo inmediato.

## **Intencionalidad como rasgo de lo mental**

Es un hecho curioso que cuando tenemos estados mentales tales como deseos, creencias, miedos, esperanzas, etc. éstos deban *ser* deseos, creencias, miedos, esperanzas etc. *acerca de* algo. Simplemente no es concebible que tengamos –por ejemplo- creencias sin una situación (real o ficticia) asociada a ella. Yo puedo creer que la tierra es cuadrada, o bien puedo creer que Lagos es el presidente de Chile, pero cualquiera sea mi creencia ésta será una creencia que (*acerca de*) algo. Acerca de lo cual es una creencia, deseo, miedo, etc. es lo que se conoce como su contenido. La propiedad de los estados mentales de poseer contenido, o bien ser *acerca de* algo, es lo que se conoce como intencionalidad.

La noción de intencionalidad también se aplica para el caso de los conceptos. Es un hecho difícilmente discutible que mi concepto de perro guarda alguna relación extraña con los perros. Esta propiedad de mi concepto de perro de estar dirigido a los perros es lo que cotidianamente referimos como la propiedad de *ser acerca de* perros y es expresión de la misma propiedad antes referida para los estados mentales: la intencionalidad.

En general, podemos decir que lo mental presenta una característica especial de estar dirigido a objetos o situaciones del mundo. Más aun, hay dos hechos muy sorprendentes en relación con todo esto. El primero es que esta característica resulta ser esencial a las entidades mentales, pues si mi concepto de perro dejara de *ser acerca de* perro dejaría de ser el concepto que es. El segundo es que esta propiedad parece ser exclusiva de las entidades mentales, siendo el caso que otras entidades no-mentales (e.g. una mesa) no *son acerca de* nada sino que simplemente son. En otras palabras, las y sólo las entidades mentales exhiben intencionalidad.

La identificación entre lo mental y lo intencional fue introducida por Franz Brentano (1874) a mediados del siglo XIX. No deja de ser llamativo, sin embargo, que Brentano haya usado la caracterización de los estados mentales en términos de su intencionalidad como argumento a favor de una ontología dualista. Y no porque parezca raro que Brentano no creyera en la reductibilidad de lo mental a lo material sino porque ignoró otro tipo de argumentos para sostener este punto. De hecho, me parece plausible pensar que ya en aquellos tiempos argumentos dualistas como los de los escéptico-solipsistas o como la ausencia de propiedades sensoriales en los estados mentales estaban empezando a caer en descrédito para algunos pensadores.

Como sea, deliberadamente Brentano puso orden en la discusión del problema Mente/Cuerpo, pero involuntariamente tendió una mano a los futuros defensores de la reductibilidad de lo mental a lo material. La razón es que aun para éstos resultaba difícil rescatar la singularidad de lo mental como fenómeno material. Si bien postulaban que los fenómenos mentales eran un tipo de fenómeno material era sumamente embarazoso no poder distinguir aquellos fenómenos de otros como –digamos- la caída de una piedra por la ladera de un cerro. La adscripción de intencionalidad a los fenómenos mentales sacó a los defensores de la reductibilidad mente-materia de este apremio.

Pues bien, considerando la ineficacia de los argumentos en contra de una postura materialista expuestos en las secciones precedentes, el problema de reducir los fenómenos mentales a otros materiales puede ser formulado del siguiente modo:

A) Un estado mental en un hombre es un estado cerebral<sup>4</sup> poseedor de una propiedad especial; a saber, su intencionalidad.

Restringimos aquí la formulación a los seres humanos por la sola razón de respetar el modo como fue planteado el problema Mente/Cuerpo en el presente capítulo, sin perjuicio de que ésta puede ser extendida a toda clase de organismos poseedores de vida mental.

Con todo, así planteado el problema lo que restaría para mostrar que no se requiere más que la postulación de un dominio ontológico (viz. el material) para explicar la fenomenología humana es explicar la intencionalidad en términos materialistas, es decir, compatibles con una descripción fisicalista del mundo. Este proyecto se conoce como el de naturalización de la intencionalidad, el cual no parece ser nada de sencillo siendo prueba de ello que durante el siglo XX numerosos intentos fueron llevados a cabo sin que sus resultados hayan conseguido el consenso necesario.

Aun así, convengamos que hay dos programas posibles para tan atemorizante tarea. El primero consiste en tratar de aislar alguna propiedad de los estados mentales *qua* estados materiales al nivel físico-químico-biológico que sea peculiar a ellos de modo que sirva como base para explicar su intencionalidad (lo que Pierce llamaría el *grounding*). Ésta sería sin dudas la apuesta mayor aun cuando también la más difícil pues, como observa Fodor (1987), pareciera que la intencionalidad simplemente no llega tan lejos en la estructura metafísica del mundo. El segundo programa posible es, en lugar de aislar alguna propiedad de los estados mentales, integrar la noción de intencionalidad a todo tipo de fenómenos mentales o no. De este modo la intencionalidad sería una propiedad –digamos- ordinaria en el mundo físico y no exclusiva de los estados mentales. Intentos como este pueden encontrarse en teorías como la

---

<sup>4</sup> Es importante observar que esto no conlleva compromiso alguno con la Teoría de la Identidad entre estados mentales y estados cerebrales toda vez que ésta exige no sólo que los primeros sean iguales a los segundos sino que esta identidad sea necesaria. Decir, en cambio, que los estados mentales son en los hombres contingentemente iguales a los estados cerebrales es simplemente trivial para quienes creen en la reductibilidad mente-materia.

biosemántica de Millikan o la semántica informacional de Dretske y serán discutidos en el capítulo 2. De momento bástenos decir que éstos suponen desperdiciar el aporte de Brentano en la caracterización de los estados mentales como poseedores de intencionalidad. En otras palabras, para teorías como éstas sigue siendo embarazoso no poder distinguir entre un estado mental y la caída de una piedra por la ladera de un cerro.

Con todas las dificultades que esto pueda involucrar, en la presente investigación intentaremos aportar una teoría de la naturalización de la intencionalidad siguiendo el primero de los programas anteriores. A continuación, en el capítulo 2, estudiaremos algunas de las teorías más prestigiosas que se han formulado para enfrentar el problema de la naturalización de la intencionalidad. Más específicamente nos ocuparemos de analizar y esbozar algunas críticas de las semánticas causales, la semántica informacional y la biosemántica –también llamada teleosemántica. Como observará el lector dejaremos fuera de dicho capítulo a las semánticas del rol conceptual, las cuales han gozado de gran prestigio durante los últimos 20 años. Sin embargo, nos ocuparemos de ellas y las dificultades teóricas que han debido enfrentar en el capítulo 5 cuando realicemos una comparación acuciosa entre éstas y la teoría de la intencionalidad que intentamos formular en el presente trabajo.

El capítulo 3, por su parte, dará comienzo a la exposición de una nueva propuesta teórica para enfrentar el problema de la naturalización de la intencionalidad. Aquí discutiremos los aspectos ontológicos y epistemológicos fundamentales asociados con la propuesta que intentaremos sostener. Esto, claro está, supondrá adelantar parte de la metodología que ocuparemos para desarrollar nuestra teoría, la cual estará determinada por niveles: uno de fenómenos biológico (y, a la vez, teleológico) y el otro netamente intencionalidad. Dicho sea de paso, aquí nos pronunciaremos por una alternativa de naturalización de la intencionalidad, que la conciba como la característica fundamental y exclusiva de ciertos fenómenos materiales llamados mentales. En consecuencia, nos alejaremos de posiciones como las de Millikan o Dretske para quienes la intencionalidad es una propiedad presente en todo tipo de entidades mentales o no.

En el capítulo 4 comenzaremos el desarrollo mismo de nuestra propuesta, partiendo por una discusión de los fundamentos de la teoría de la mente a la cual adheriremos: el funcionalismo. Luego, estableceremos nuestro punto de partida en la teoría de la Autopoiesis de Maturana y Varela, la cual nos entrega una concepción materialista del fenómeno de la vida. Desde allí intentaremos construir una concepción materialista de los fenómenos mentales en tanto

exclusivos de los entes vivos. Introduciremos para esto la noción de psicopoiesis, la cual será un tipo de proceso autopoiético asociado a lo mental. Finalmente, mostraremos cómo un funcionalismo teleológico respecto de la mente como el que proponemos puede ayudar a solucionar una serie de problemas teóricos considerados insolubles para otras versiones de las teorías funcionalistas.

Finalmente, en el capítulo 5 nos apoyaremos de la noción de psicopoiesis del capítulo anterior para desarrollar una noción materialista de la orientación (en sentido genérico) al mundo de lo mental. Esto es lo que llamaremos la especificidad de los procesos psicopoiéticos, la cual estará definida en términos de la conservación o no conservación de una cierta estructura de funcionamiento propio de los organismos vivos dotados de mente. Luego, aplicaremos esta idea de la orientación de lo mental al mundo para un tipo particular de entidad mental como son los conceptos y en general, cualquier tipo de representación mental. El resultado será una propuesta de explicación de la intencionalidad de lo mental y del contenido de las representaciones. Para concluir, desarrollaremos una comparación de nuestra propuesta con otra relativamente cercana como son las semánticas del rol conceptual. De este modo, someteremos nuestra teoría a las críticas más devastadoras que han debido enfrentar dichas semánticas, observando que ninguna de ellas resulta dañina para ella.

## Capítulo 2

El inductivismo comenzó a ser seriamente cuestionado en la Epistemología del siglo XX. Desde entonces, se cree que hacer conocimiento es en gran medida una cuestión de imaginación. Nuestro acceso a la realidad es indudablemente limitado, razón por la cual parte de hacer conocimiento es imaginar cómo funciona el mundo. La labor de los científicos, historiadores, filósofos, etc. consiste comúnmente en formular teorías y mecanismos subyacentes a la realidad visible. Pero, naturalmente, para que este procedimiento tenga valor explicativo, aquellos deben estar sujetos a ciertas restricciones. De otro modo, todos los postulados tendrían el mismo mérito y no habría manera de decidir entre dos imaginaciones del mismo trozo de realidad.

Conforme con esto, resulta de gran utilidad al evaluar una teoría estipular previamente y con claridad las restricciones que ésta debe satisfacer. Así, dado que el propósito del presente capítulo es estudiar algunas de las teorías filosóficas más célebres en oferta para explicar la semántica de los estados mentales, comenzaremos nuestra tarea definiendo un marco mínimo que cualquier semántica debe respetar. A continuación, someteremos cada una de ellas a este mínimo exigible, donde nos encontraremos con el lamentable sorpresa de que ninguna consigue pasar la prueba de blancura. La conclusión será, pues, evidente: habrá que imaginar nuevas teorías del contenido intencional.

### **El marco mínimo**

Como ya debe ser claro, una semántica filosófica es una teoría que intenta explicar la relación entre un símbolo S y un objeto O cotidianamente referida como “S es acerca de O”. Sin embargo, aun se discute entre los filósofos si la propiedad de “ser acerca de” es intrínseca de toda clase de entidades (mentales o no) o si, por el contrario, ésta es patrimonio exclusivo de las entidades mentales (conceptos, estados mentales, etc.) mientras que las restantes entidades sólo la tomarían prestada de aquéllas. Quienes se inclinan por esta última opción han agregado a la intencionalidad de las entidades mentales el adjetivo “originaria” como forma de distinguirla de la de otras entidades no-mentales. Para independizar nuestro marco para las semánticas filosóficas de esta polémica, exigiremos a toda semántica filosófica que explique

al menos la intencionalidad originaria; esto es, la capacidad de las entidades mentales de *ser acerca de* algo, sin exigir que lo hagan de otras entidades.

Así puesto el problema, tenemos, en primer lugar, un conjunto de restricciones exigibles a cualquier teoría –semántica o no- que tienen relación con su estructura interna. Entre ellas se cuentan la consistencia lógica, la plausibilidad de sus deducciones y, en general, todos los aspectos que determinan su coherencia. Habitualmente los filósofos ponen especial atención a este tipo de consideraciones y, de este modo, son satisfechas por la inmensa mayoría de las semánticas en ofertas. En consecuencia, daremos mayor importancia a un segundo tipo de restricciones que llamaremos externas<sup>5</sup>, sin perjuicio de que eventualmente alguna de las semánticas filosóficas pueda incurrir indeseadamente en el desacato de alguna de las conocidas restricciones de coherencia interna.

Entenderemos por restricciones externas todos aquellos constreñimientos que enfrenta una teoría en tanto se inserta en un ámbito teórico más amplio del cual participa. Éstos van desde la credibilidad que ofrece estar armonía con otras teorías circundantes hasta la obligación de respetar ciertos principios ontológicos fundamentales. Entre estos últimos, podemos contar el supuesto materialista en las ciencias naturales, cuestión que se ha venido extendiendo como exigencia a la filosofía analítica desde los trabajos del Círculo de Viena. Hoy por hoy resulta un requerimiento para cualquier teoría de la mente que pretenda ser tomada en consideración evitar cualquier compromiso con posiciones metafísicas dualistas. Esta es la razón por la que frecuentemente oímos hablar en este ámbito teórico de reduccionismo y eliminativismo y rara vez de dualismo o idealismo. Esta es también la razón por la que la adscripción a estas últimas funciona como argumento en contra de ciertas posiciones respecto del significado (Fodor y Lepore 1992).

Pues bien, para cumplir con nuestro compromiso de proponer un marco mínimo de restricciones para las semánticas filosóficas, nos concentraremos en dos restricciones en lo que resta. Estas serán las siguientes:

---

<sup>5</sup> Al hacer esta taxonomía me he inspirado con mayor o menor apego a las ideas de Larry Laudan (1981).

1) *Restricción Semántica*: Cualquier explicación de las propiedades semánticas de las entidades mentales no debe incluir dentro de su vocabulario referencias –directas o indirectas– a propiedades semánticas.

2) *Restricción Epistémica*: Cualquier explicación de las propiedades semánticas de las entidades mentales no debe incluir dentro de su vocabulario referencias –directas o indirectas– a capacidades epistémicas de los individuos.

La primera de ellas es, sin dudas, una restricción interna; esto es, relevante para garantizar la coherencia interna de la teoría. La razón es que si lo que se quiere explicar son las propiedades semánticas de las entidades mentales, éstas no pueden ser supuestas por la teoría. Suponer un fenómeno como existente no es explicarlo. Cualquier teoría que quiera explicar un fenómeno debe cuando menos hacerlo en términos de otros fenómenos. En caso contrario, no explicaría nada.

Sin embargo, no siempre es fácil cumplir con la Restricción Semántica. Esto probablemente se deba a que en nuestra cultura separamos fuertemente dos ámbitos ontológicos: el material y el mental. Las propiedades que los filósofos comúnmente llaman semánticas serían, según lo discutido en el capítulo 1, características de este último ámbito. Pero, a pesar de esta separación cultural, en nuestra vida cotidiana ambas dimensiones se ven constantemente entrelazadas, haciendo que ocasionalmente sea muy fácil confundir una referencia a propiedades semánticas con otra a propiedades físicas.

Esto también explica por que el problema Mente/Cuerpo ha sido tan indómito durante tanto siglos. Si fuese tan fácil separar las propiedades semánticas de las no-semánticas entonces sería presumiblemente accesible postular la relación –si alguna– existente entre ellas. Sin embargo, no sólo esto último no es tarea sencilla sino que incluso caracterizar filosóficamente las propiedades semánticas ha tomado muchas generaciones de pensadores. Baste observar que hasta mediados del siglo pasado aun persistían discusiones respecto de cuál era el significado de una palabra; si este era un colgador de descripciones (Russell 1918), si su referencia (Wittgenstein 1921) o sus usos (Strawson 1950).

Como sea, es característico de este tipo de restricciones, llamadas internas, el hecho de que actúan *a priori*; esto es, se aplican a cualquier teoría, sólo en virtud de su coherencia y valor explicativo, independientemente de cuáles teorías complementarias sean aceptadas en una cierta comunidad. La segunda restricción, en cambio, llamada Restricción Epistémica, nace de

la necesidad de situar cualquier semántica en armonía con otras teorías filosóficas adyacentes y depende, pues, de la información *a posteriori* deducida de aquello que la comunidad filosófica acepta como cierto.

Pues bien, dos son los postulados teóricos que abogan por la necesidad de respetar la Restricción Epistémica. Éstos son los siguientes:

Postulado 1: Las propiedades semánticas de las entidades mentales poseen existencia metafísica.

Con esto queremos decir que la relación entre, digamos, un concepto y su referente no es meramente ilusoria, sino que las propiedades semánticas poseen existencia metafísica. Este supuesto está en la base de cualquier intento de naturalizar la intencionalidad, pues difícilmente alguien emprendería un proyecto tal si no creyese que hay algo en la composición metafísica del mundo que explicar en términos de las ciencias naturales.

Sólo podrían eximirse de este postulado quienes no creen en la dimensión semántica de la mente. Entre estos encontramos filósofos como Stich, quienes creen que el solo aspecto sintáctico es suficiente para explicar todos los fenómenos mentales. También los Conductistas Radicales, quienes refutan incluso la existencia de la mente como objeto de estudio. Para todos ellos el mero problema del significado y las propiedades semánticas carece de sentido.

Más aun, si entendemos el problema del significado desde la perspectiva de la necesidad de una ciencia psicológica, acaso también podríamos excusar de adherir al Postulado 1 a algunas variantes de teorías instrumentalistas respecto de las actitudes proposicionales. En efecto, si la necesidad de apelar al contenido proviene de la necesidad de individuar las actitudes proposicionales por medio de éste, entonces quienes no asignan realidad metafísica a las actitudes bien pueden no asignársela al contenido de éstas. Para filósofos instrumentalistas como Dennett (1981) la intencionalidad es parte de un discurso explicativo más o menos eficaz que no necesariamente tiene un correlato metafísico. La psicología popular y la postulación de intencionalidad en los estados mentales serían, pues, recursos momentáneos a la espera de una teoría mejor y, por lo tanto, no tendría sentido intentar su naturalización. Como sea, con o sin una ciencia psicológica, aun persiste la intuición evidencial de que nuestros pensamientos *son acerca de* porciones de mundo. En principio al menos, uno podría intentar explicar esta intuición sin tener en la mira una teoría de la conducta.

Para delimitar medianamente el campo de acción del postulado anterior, convengamos que todos quienes son realistas epistemológicos y creen en la psicología popular como la mejor explicación de la conducta deben aceptarlo. La razón es que si uno cree en la realidad de las actitudes, entonces debe aceptar la individuación por medio de su contenido y, por lo tanto, la realidad de las propiedades semánticas. Otra opción como, por ejemplo, creer en la individuación funcionalista de las actitudes (i.e. la individuación funcionalista de la *creencia-que-P*) conduce inexorablemente al holismo (Fodor 1987). Y, de allí, hay un solo paso al idealismo. Pero no se puede ser idealista metafísico y al mismo tiempo realista epistemológico. Si uno no supone que nuestros conceptos están determinados de algún modo por la realidad externa difícilmente podría uno aceptar que las teorías hechas a partir de dichos conceptos mientan entidades o propiedades reales. Ser partidario del realismo epistemológico supone ser antes partidario del realismo metafísico.

Pues bien, si las propiedades semánticas poseen existencia metafísica y no constituyen un efecto ilusorio entonces éstas pueden ser explicadas con independencia de la estructura de creencias de los individuos. Luego, la intencionalidad debe ser entendida como una propiedad presente en el mundo antes que circunscrita al modo como interpretamos el mundo y, por lo tanto, las capacidades epistémicas de los sujetos cognoscentes no deben jugar rol alguno en su explicación. La Restricción Epistémica es, pues, una forma de poner en armonía una teoría de la intencionalidad con el Postulado 1.

Postulado 2: Una teoría de las capacidades epistémicas requiere previamente la existencia de propiedades semánticas.

El principal argumento que soporta este postulado es extraordinariamente sencillo: la intencionalidad es ontológica y explicativamente previa a las capacidades epistémicas. Si es que los hombres tenemos la capacidad de hacer teorías es porque antes tenemos la capacidad de tener pensamiento *acerca de* algo. Si hacer teorías es de un modo u otro interpretar (explicar) el mundo entonces ¿cómo podríamos interpretar algo *acerca de* lo cual no podemos tener pensamientos? Nuestra capacidad de realizar hipótesis, formular explicaciones, etc. requiere de la característica de nuestras mentes de instanciar estados que poseen propiedades semánticas y, en definitiva, intencionalidad. No hay capacidades epistémicas sin una conexión semántica con el mundo. De otro modo, la teoría atómica no podría *ser acerca de* los átomos, ni la biología celular *acerca de* las células.

Otra manera estándar en que se ha formulado el mismo argumento es que una metafísica antecede lógicamente a una epistemología. En efecto, si lo que se quiere es hacer una teoría de la forma como los sujetos conocen el mundo, es una necesidad lógica establecer previamente qué es el mundo objeto de conocimiento, por una parte; y, la propiedades de las mentes de relacionarse con éste, por otra. De este modo, salvo posiciones idealistas, cualquier teoría epistemológica que adhiera al realismo en algún grado requiere de una explicación independiente de las propiedades semánticas de las entidades mentales.

Pero si esto no bastara, existe un tercer argumento para defender el Postulado 2 construido a partir de lo que se ha denominado evaluación semántica de los estados mentales. Si, como se suele decir, las creencias (un tipo de estado mental) son susceptibles de evaluación semántica —esto es, en términos de verdaderas o falsas— las propiedades semánticas de un estado mental (e.g. una creencia) están estrechamente vinculadas a las condiciones de verdad del mismo. Esto es evidente y ha sido ampliamente aceptado en la filosofía de la mente. Luego, lo que se tenga para las condiciones de verdad plausiblemente se tendrá también para las propiedades semánticas. Pero es un hecho innegable que lo mínimo que se le puede pedir a una epistemología razonable (excluyendo al constructivismo u otras análogas) es que las condiciones de verdad de las representaciones mentales no dependan de las capacidades epistémicas de los individuos en que éstas se instancian. En caso contrario, tendríamos que una misma creencia instanciada en dos individuos con distintas capacidades epistémicas podría ser falsa en un caso y verdadera en el otro. Esto es lo que se conoce como relativismo respecto de la verdad, cuestión que deja sin explicación muchos fenómenos epistémicos como, por ejemplo, que los debates científicos sean productivos.

Ahora bien, si para hacer una epistemología razonable necesitamos condiciones de verdad para los enunciados que no dependan de las capacidades epistémicas de los individuos y dichas condiciones de verdad son una expresión de las propiedades semánticas, entonces estas últimas deben también ser independientes de las capacidades epistémicas de los individuos. Esto es precisamente lo que plantea el segundo postulado.

Con todo, sean cuales sean los argumentos, siempre habrá una salida para no aceptar el Postulado 2; ésta es, naturalmente, el idealismo o, acaso, su versión contemporánea, el constructivismo. Desdichadamente, sin embargo, todas ellas están fuera de discusión en el ámbito de la filosofía de la mente debido a que ésta participa en la ciencia cognitiva. El sustantivo “ciencia” prohíbe prestarles cualquier consideración.

Como sea, el postulado 2 sienta una condición aun más fuerte sobre el carácter ontológico de las propiedades semánticas que la estipulada en el postulado 1, cuya importancia quedará de manifiesto cuando estudiemos más abajo las semánticas causales e informacionales. Ésta es que no sólo la intencionalidad existe como propiedad de los estados mentales con independencia de la existencia de capacidades epistémicas en los individuos sino que además ésta es lo que llamaremos una propiedad metafísica. Hagamos un poco de metafísica para explicar qué entenderemos por “propiedad metafísica”. Consideremos los siguientes dos ejemplos. Primero, la velocidad promedio de un conjunto de partículas en movimiento puede considerarse sin mayor inconveniente una propiedad de dicho sistema. Incluso podríamos decir, en una ontología bastante permisiva, que ésta existe con independencia de nuestra capacidad de hacer matemáticas y estadística: simplemente está allí nos demos cuenta o no de ello. Lo mismo podríamos decir de una segunda propiedad como la energía del mismo sistema de partículas. Sin embargo, estas dos propiedades difieren en un aspecto ontológico fundamental; a saber, la última supone poderes causales para el sistema en ciertas situaciones posibles mientras que la primera no.

La energía de un cierto sistema de partículas puede determinar, por ejemplo, las consecuencias de una colisión de éste con otro conjunto de partículas. Más aun, esto nada tiene que ver con nuestra interpretación del fenómeno. La velocidad promedio de las partículas del sistema, en cambio, puede ser útil en el análisis del mismo sólo en la medida en que poseemos una teoría estadística que nos permita relacionar dicha propiedad con su energía. En general, si es que las propiedades estadísticas de un sistema físico son señal de la potencia causal de un sistema de partículas esto se debe a su eficacia como instrumento de cálculo aproximado de ciertas propiedades del sistema que sí poseen potencia causal. Su utilidad se deriva, en todo caso, de nuestra capacidad para formular buenas teorías matemáticas que aproximen la realidad y, en definitiva, de nuestras capacidades epistémicas. De este modo, basándonos en esta distinción, llamaremos propiedades metafísicas a aquéllas que suponen poderes causales en los objetos en que se instancian independientemente de nuestras teorías y capacidades epistémicas en general. Así, la energía de un sistema de partículas caerá bajo esta denominación y no así la velocidad promedio del mismo.

El mérito de esta distinción se encuentra en el hecho de que si bien propiedades no-metafísicas como la velocidad promedio gracias a que fueron ideadas para ser independientes de nuestra estructura de creencias pueden considerarse existentes en una ontología suficientemente permisiva, éstas siguen siendo simples herramientas cognitivas. En

consecuencia, éstas no pueden figurar en la comprensión de la estructura metafísica última del mundo pues no juegan ningún rol en lo que ocurra en él en tanto no suponen potencia causal alguna. Hacer metafísica es, pues, hacer teorías que expliquen lo que ocurre en el mundo sin acudir a propiedades que no sean propiedades metafísicas. Otros esfuerzos teóricos como, por ejemplo, la introducción de nociones estadísticas en el análisis de los fenómenos físicos (viz. la llamada Física Estadística) constituyen potentes herramientas predictivas pero no explicaciones de la realidad metafísica última. Sólo en virtud de la presencia de sujetos con intereses cognitivos estas teorías no-metafísicas poseen relevancia y ésta, en todo caso, no va más allá de su rol meramente práctico.

Ahora bien, es un hecho difícilmente cuestionable que la intencionalidad es parte de la estructura metafísica del mundo, lo que corresponde a decir que ésta es una propiedad metafísica. Esto se debe a que para una entidad mental poseer intencionalidad supone poderes causales en situaciones posibles. Es evidente que la instanciación de mi concepto PERRO suele causar conductas (e.g. respuestas motrices) que no causaría si ésta no fuera *acerca de* perros (viz. si fuera *acerca de* tigres o simplemente no fuera *acerca de* nada). Consecuentemente, si la intencionalidad pertenece a la estructura metafísica del mundo ésta debe poder explicarse en términos que prescindan de propiedades no-metafísicas (como la velocidad promedio de cierto sistema de partículas). Una explicación que no respeta este requisito si bien no necesariamente será incorrecta, no pasará de ser un mero instrumento práctico que no ofrecerá un aporte sustantivo a la comprensión de tan importante problema.

Pero eso no es todo. Si el objetivo de formular una teoría naturalista de la intencionalidad es avanzar en la comprensión de las capacidades cognitivas, como sugiere el postulado 2, entonces más vale que lo hagamos de modo que nos permita su comprensión como parte del entramado metafísico del mundo. Las teorías que acuden a propiedades no-metafísicas si bien pueden tener valor práctico tienen también la indeseable característica de ser débiles fundamentos sobre los cuales construir otras teorías. No es un secreto que a pesar de la capacidad de la Física Estadística para explicar el comportamiento de ciertos fluidos, cuando fue necesario comprender otros fenómenos aledaños (e.g. los plasmas, los fenómenos caóticos, etc.) ésta no sirvió como punto de partida. Aparecieron así teorías independientes como la física no-lineal, la mecánica cuántica etc., las cuales intentaban competir en el ámbito de explicación metafísica.

En resumen, toda semántica estará impedida de recurrir a propiedades no-metafísicas bajo pena de constituir un mero instrumento práctico-cognitivo en lugar de una explicación metafísica última. Respetar este orden explicativo será un requisito a tener en cuenta si se pretende estar en armonía con la Restricción Epistémica.

## **Semánticas Causales**

Después del auge de la semántica del rol conceptual nacen en Wisconsin las que más tarde serían el último grito de la moda: las semánticas causales. Supuesta una interpretación matemática de la noción de función, los partidarios del funcionalismo terminaron por reconocer que intentar individuar el contenido de las actitudes funcionalmente conduce inexorablemente al holismo. Y, salvo excepciones, nadie quiere comprometerse con el holismo, porque del holismo al idealismo hay un pequeño paso y el idealismo es un precipicio filosófico. Suponer que lo que ocurre en nuestra mente es independiente del mundo exterior (si es que existe tal cosa) invalida nuestra posibilidad de hacer teorías, formular hipótesis y, en general, desarrollar conocimiento del mismo. Acaso invalidaría también la individuación funcional del contenido, raíz de todo este problema...

Una versión primitiva aunque representativa de la idea fundamental de las semánticas causales diría lo siguiente:

V1) Una entidad mental es acerca de un objeto si, y sólo si, la primera es causada por el último.

Consecuentemente, mi concepto de María sería acerca de María en la medida que es causado por ella; aun cuando la cadena causal que conecta a ésta última con mi concepto puede ser bastante larga y contener muchos eslabones intermedios.

Claro está, esta versión de las semánticas causales difícilmente pueda ser encontrada en la literatura del ámbito de la filosofía de la mente, pues se trata de una extremadamente cruda y sujeta a un sinnúmero de dificultades evidentes, algunas de las cuales repasaremos someramente a continuación. De momento bástenos constatar dos hechos. Primero, que si bien pocos osarían defender una versión como ésta, ella es perfectamente clasificable como una semántica causal, pues es la mera idea fundamental sobre la cual los filósofos asiduos a estas doctrinas (Devitt 1981, Stampe 1977, Fodor 1987) han elaborado las suyas enmendándola de modo de evitar ciertas dificultades. Segundo, que (V1) es una versión de las semánticas causales

clasificable como instanciación/instanciación (token/token) pues establece una relación entre tokens para explicar la intencionalidad.

Pues bien, la primera crítica que se podría esgrimir contra (V1) es que en su formulación no se logra distinguir cuáles son los eslabones relevantes en la cadena causal para efectos de la intencionalidad. Así, si, por ejemplo, María causó en Juan el concepto MARÍA<sup>6</sup>, y este, a su vez, causó en mí el concepto MARÍA relatándome las circunstancias en que conoció a esta mujer, (V1) no ofrece manera de decidir si mi concepto MARÍA es *acerca de* Juan o *de* María, en tanto que ambos son dos eslabones en la misma cadena causal. Consecuentemente, siguiendo (V1) rigurosamente, mi concepto *de* María sería también *acerca de* Juan, cuestión lisa y llanamente intolerable.

En segundo lugar está el problema de los conceptos sin extensión. Como fue formulada, (V1) no tiene ninguna posibilidad de explicar las propiedades semánticas de ciertos conceptos cuyos referentes sabemos que no existen. Un ejemplo de este tipo de concepto sería el de unicornio. En tanto no hay unicornios en el mundo real, no puede ser el caso que los conceptos UNICORNIO signifiquen lo que significan en virtud de haber sido causados por unicornios. Más bien, esta formulación primaria de las semánticas causales pareciera indicar que dicho concepto no posee significado o, si lo posee, es de naturaleza distinta del de mi concepto PERRO. Ninguna de estas alternativas puede dejar a alguien satisfecho.

En tercer lugar tenemos el problema del error. Esto es, la idea es que en ocasiones los individuos se equivocan e instancian la representación mental CABALLO en presencia de vacas. Presumiblemente será esta última la que cause el concepto en la cabeza del sujeto. Pero, entonces, ¿significa esto que el concepto CABALLO instanciado *es acerca de* vaca? ¿Significa esto que una persona que frente a una vaca erróneamente invoca pensamientos *acerca del* caballo de Napoleón súbitamente está pensando *acerca de* la vaca de Napoleón? Evidentemente que no.

Con todo, como advertíamos más arriba, las dificultades hasta aquí observadas de esta versión primitiva de las semánticas causales han sido solucionadas con mayor o menor éxito

---

<sup>6</sup> Adoptaremos la convención estipulada en *Concepts: Where Cognitive Science were wrong*, de Fodor, para referirnos a la representación mental (PEDRO), a la expresión del lenguaje ('Pedro') y al objeto referido (Pedro).

realizando enmiendas y reformulaciones de la misma. Más aun, estas reformulaciones siguen conservando de un modo u otro el mismo principio explicativo, en tanto que se basan en la relación causal para dar cuenta de la intencionalidad de las representaciones mentales. Sin embargo, hay una objeción en contra de (V1) que jamás podrá ser enmendada, puesto que hacerlo supondría renunciar *ipso facto* a la relación causal como principio explicativo. Ésta es que (V1) no es una explicación de la intencionalidad eminentemente metafísica y, por lo tanto, no sienta bien a la Restricción Epistémica.

En efecto, la historia –causal u otra- no es una propiedad metafísica y, por lo tanto, no puede participar de una explicación metafísica de la intencionalidad. Consideremos un ejemplo para ilustrar este hecho. Una partícula P situada en un cierto lugar del espacio en un cierto instante, tiene entre sus propiedades metafísicas su velocidad, su masa, su carga eléctrica, etc. Éstas son las propiedades relevantes para efectos de saber cómo esta partícula participa del entramado causal del mundo. Cómo dicha partícula llegó a tener la velocidad, la masa y la carga que tiene no aporta nada a su potencia causal que no haya estado antes incluido entre sus propiedades metafísicas y, por lo tanto, no posee ninguna relevancia en una descripción metafísica del mundo. Sólo en la medida en que hay sujetos interesados en saber qué hace que una partícula cualquiera llegue a poseer dichas propiedades metafísicas, la historia de la partícula adquiere importancia.

Paradójicamente, por el mismo argumento, la historia causal de una entidad mental no supone poder causal alguno para ésta. Sin embargo, reitero (ver arriba), la intencionalidad (*a fortiori*, las propiedades semánticas) debe ser una propiedad metafísica de toda entidad mental que la posea, esto es, una que supone poderes causales para ella –y, por lo tanto, relevancia en la estructura metafísica. Luego, una explicación metafísica de la intencionalidad no puede depender de la historia causal del símbolo, pues ésta juega un rol sólo en un contexto práctico-cognitivo, donde hay individuos interesados en predecir. La apelación a la historia causal de un símbolo es, pues, una desconsideración de la Restricción Epistémica antes discutida.

Como sea, existe la intuición de que la historia de hecho causa conductas en nosotros, seres humanos, lo cual justificaría entenderla como propiedad metafísica. Alguien podría decir, por ejemplo, que una historia de fracasos en cocinar arroz en una sartén causará que en el futuro lo intentemos con una olla. Más generalmente, pareciera que la potencia causal de la historia posibilita el aprendizaje en los seres humanos. Sin embargo, esto no es estrictamente así. La réplica a una aseveración tal sería que el cambio en nuestras conductas no es causado por la

historia sino que por estados mentales que son *acerca de* la historia, lo cual es muy distinto. Para recalcar este hecho consideremos un sencillo experimento filosófico. Imaginemos cómo funcionaba el mundo antes que hubiera en él organismos dotados de mente (*a fortiori*, sin capacidades epistémicas). Una cuestión es evidente: en un mundo así no habría aprendizaje. Más aun, según lo que estamos tratando de argumentar, todo lo que pueda ocurrir en él está completamente determinado por las propiedades metafísicas del estado inmediatamente anterior. Si existiera un dios pasivo y omnipresente, que conociera cada propiedad metafísica de cada partícula en un cierto estado de este mundo, éste no necesitaría recurrir a los estados anteriores para saber cuál será el estado siguiente. El pasado no es relevante para efectos de determinar qué ocurrirá en el futuro. Consecuentemente, la historia no es una propiedad metafísica. Luego, si la estructura metafísica última del mundo no ha cambiado desde los tiempos en que no habían seres inteligentes la historia no puede servir como base para una explicación metafísica de la intencionalidad.

Para ver hasta qué punto las semánticas causales fracasan en su afán de aportar una teoría metafísica de la intencionalidad, estudiemos, a continuación, una de sus mejores formulaciones, como es aquella desarrollada por Fodor (1987). Ésta plantea una innovación fundamental respecto de (V1) que consiste en postular que la relación semántica es propia de los tipos y no de las instancias (Von Eckardt 1995). Así, un concepto-tipo *sería acerca de* una propiedad cuando instancias de esta última *confiablemente* causan instancias del primero. Aquí, el adverbio “confiablemente” introduce una condición modal a la anterior formulación. Traducido a la interpretación kripkeana de las modalidades diría:

(V2) Un concepto-tipo ‘A’ denota una propiedad-tipo A cuando en todo mundo nomológicamente cercano al nuestro ejemplares de A causan instancias de ‘A’<sup>7</sup>.

Así, por ejemplo, en esta formulación el problema del error antes mencionado se transforma en el de la disyunción, pues el hecho de que una vaca cause instancias del concepto CABALLO no supone que dicho concepto deje de *ser acerca de* caballos. Esto debido a que la intencionalidad es aquí una propiedad de los conceptos-tipos y no de sus instancias. Aun

---

<sup>7</sup> El mismo Fodor sugiere que V2 puede corresponder a algún tipo de relación nomológica (legaliforme) entre estados. Sin embargo, aquí no haremos caso de esta observación pues trataremos esta idea en profundidad en la sección siguiente cuando estudiemos la semántica informacional.

así, sabemos que de vez en cuando ejemplares de vacas causan instanciaciones del concepto CABALLO, sin perjuicio de que habitualmente sean ejemplares de caballos las que causen instanciaciones del concepto CABALLO. En consecuencia, tenemos que ejemplares de *caballos o vacas* causan instanciaciones del concepto CABALLO, cuestión que reduce el problema a averiguar qué hay de distinto en las instanciaciones de CABALLO causadas por caballos de aquéllas causadas por vacas. Este es el problema de la disyunción, el cual resulta mucho más abordable para esta formulación de las semánticas causales de lo que era el del error para (V1). De hecho, el propio Fodor ofrece una solución más que aceptable conocida como la dependencia asimétrica.

Pero dejemos de lado este problema suficientemente discutido para concentrarnos en el argumento que hemos tratado de desarrollar en contra de las semánticas causales. Como decíamos anteriormente ninguna versión de las semánticas causales podrá desembarazarse jamás del hecho que la historia causal no es una propiedad metafísica. Consecuentemente, si es que (V2) constituye una buena teoría de la intencionalidad pocas dudas pueden haber respecto de que esta teoría no es una teoría metafísica (en el sentido antes establecido) en tanto recurre en su explicación a la historia causal de las instanciaciones (*tokens*) de los símbolos mentales. De este modo, podríamos decir que (V2) es al estudio de los estados mentales intencionales lo que la Física Estadística al estudio de los gases. Si bien el promedio de las velocidades de las partículas que componen un cierto gas (propiedad no-metafísica) entrega una condición suficiente para que el gas tenga una cierta energía (aun cuando aproximada), este solo hecho no permite comprender qué hay en la estructura del metafísica del mundo que hace que los gases se comporten como tales. Análogamente, si bien puede que la historia causal entregue una condición suficiente para que los estados mentales posean intencionalidad esto no aporta respecto nada de qué hay en la estructura metafísica del mundo que hace que los estados mentales intencionales se comporten como tales.

Acaso el aspecto más flagrante de las dificultades de una teoría no-metafísica de la intencionalidad como (V2) es su incapacidad para dar cuenta de ciertas características indiscutibles de las entidades mentales intencionales. Por ejemplo, es un hecho ampliamente aceptado (Fodor 1998) que los conceptos poseen su contenido intencional de manera esencial; esto es, que si, por alguna razón, una instanciación mi concepto PERRO no fuera *acerca de* perro dejaría también de ser una instanciación del concepto PERRO. Por otro lado, resulta un tanto embarazoso que la historia causal de las instanciaciones de mi concepto PERRO, base explicativa de sus propiedades semánticas, no sea esencial al mismo. De hecho, aceptar el

problema de la disyunción e intentar solucionarlo supone un reconocimiento de este hecho por los partidarios de las semánticas causales. Si instancias de mi concepto PERRO pueden ser causadas por perros o por coyotes, la historia causal de éstas simplemente no es esencial a las mismas. Consecuentemente, las semánticas causales entregan una explicación de una propiedad esencial a los conceptos en términos de otra que no lo es.

Con todo, Fodor (1987) observa que el problema de la naturalizar la intencionalidad se reduce al de establecer condiciones suficientes para que un símbolo mental sea *acerca de* algo en términos que no sean intencionales ni semánticos. Y, de hecho, puede que esté en lo cierto. Puede que con el establecimiento de condiciones suficientes no-intencionales y no-semánticas baste para demostrar que hay algo en la estructura metafísica material del mundo de lo cual proviene (superviene) la intencionalidad —claro está, sin decir al mismo tiempo qué es ese algo. De ser así, pareciera que (V2) es un candidato a teoría de la naturalización de la intencionalidad. Sin embargo, como ya fue discutido, una teoría que apele a propiedades no-metafísicas como la historia causal tiene el defecto de resultar intrascendente para una comprensión metafísica del fenómeno de la intencionalidad y, por lo tanto, no es un buen punto de partida para la comprensión de otros fenómenos que dependen de ella como son, según lo establecido en el postulado 2, las capacidades epistémicas. En consecuencia, no es que debamos descartar de plano las semánticas causales, pero si podemos construir una teoría metafísica de la intencionalidad; esto es, una que recurra sólo a propiedades metafísicas en su explicación, valdría la pena hacerlo. En los capítulos 4 y 5 intentaremos llevar a cabo este proyecto.

## **Semántica Informativa**

En cierta ocasión estando con mi perro me pregunté qué hace que el movimiento de su cola signifique algo. La única respuesta que me vino a la mente fue que este fenómeno está asociado a un estado psicológico del perro de alegría o entusiasmo. La pregunta cambiaba, entonces, a ¿qué tipo de asociación existe entre los estados psicológicos de un perro y los estados de su cola? Mi respuesta fue que ni causal, ni biológica sino que estadística. Extrañamente cuando tratamos cotidianamente con perros aprendemos que frecuentemente cuando mueven su cola tienen también conductas que calificaríamos como joviales. No disponemos de ninguna teoría que conecte causalmente estos dos fenómenos. (Probablemente no haya ninguna razón biológica para que esto sea así del mismo modo como no hay ninguna

razón biológica para que el movimiento vertical de la cabeza en los seres humanos signifique aprobación.) Sin embargo, constatamos que dichos fenómenos están asociados por el simple hecho de que cuando ocurre uno (frecuentemente) también ocurre el otro.

Claro que la asociación no tiene que ser necesariamente estadística. Puede haber razones más profundas para asociar dos fenómenos. Esto es lo que ocurre cuando pensamos en por qué lo que marca un termómetro tiene significado. No es que los hombres hayan constatado empíricamente a través de su evolución que estos extraños objetos que son los termómetros marcan frecuentemente algo que cuantifica la temperatura. De hecho, durante gran parte de la vida del hombre en la tierra ni siquiera existieron termómetros. Más bien, estos son aparatos diseñados para, por medio de una ley científica, dar una medida de la temperatura. La asociación entre lo que marca un termómetro y la temperatura procede, pues, por medio de una ley.

La teoría llamada semántica informacional se basa en un principio como éste para explicar la intencionalidad. La idea deriva de la teoría de la información de Dretske (1981), según la cual, la información es un bien de intercambio perteneciente al dominio ontológico del mundo. En otras palabras, la información siempre está disponible independientemente de que hallan individuos que la utilicen alrededor. Grosso modo, un cierto estado  $E_1$  porta información sobre otro  $E_2$  cuando no es posible que se tenga  $E_1$  a menos que se tenga  $E_2$ . Así, por ejemplo, diría Dretske, si el termómetro marca 5 esto significa que la temperatura es 5 debido a que, si el termómetro está funcionando correctamente (i.e. si no hay ruido en el canal de comunicación) no es posible que marque 5 a menos que la temperatura sea 5.

Puesta de otro modo, la teoría postula que  $E_1$  porta información sobre  $E_2$  cuando existe una conexión nomológica (viz. legaliforme) entre  $E_1$  y  $E_2$ . Consecuentemente, la información sería un asunto de leyes del mundo. Esta noción más la idea de que el significado es información es lo que se conoce como semántica informacional.

Es importante destacar que la semántica informacional bien puede considerarse distinta de las semánticas causales. Esto debido a que la condición de existencia de una ley entre dos estados no supone necesariamente una conexión causal entre ellos. Por el contrario, la semántica informacional resulta de una extensión de las semánticas causales pues la conexión causal es una de las muchas conexiones legaliformes posibles. Bien podría darse el caso de que un estado  $E_1$  fuese causado por otro  $E_2$  y que, sin embargo, fuera  $E_2$  el que portara información

sobre E1, en la medida que existiese una ley que dijera que cada vez que se tiene E2 se tiene también E1. Consecuentemente, la semántica informacional indicaría una relación semántica distinta de la que indicaría una semántica causal. Según la primera, E2 significaría E1, mientras que en la segunda E1 significaría E2.

Más aun, no todas las relaciones causales entre dos estados pueden constituirse en leyes. Para que esto ocurra la relación causal debe satisfacer una condición modal adicional; esto es, que, por ejemplo, el estado E1 cause el estado E2 *en todo mundo nomológicamente cercano al nuestro*. Si un estado causa *contingentemente* otro estado entonces dicha conexión causal no constituye una ley y, por lo tanto, el segundo no porta información sobre el primero. Conforme con esto, la semántica causal de Fodor, llamada anteriormente (V2), es una bastante cercana a la semántica informacional.

Por otro lado, la noción de ley en la teoría dretsiana tiene una connotación especial. Esto debido a que si la información ha de ser parte de la estructura ontológica del mundo, entonces la noción de ley involucrada en su definición debe participar igualmente de dicha estructura. Consecuentemente, las leyes están presentes en el mundo independientemente de si nosotros nos percatamos de ello. Las leyes representan, pues, el orden en el mundo y nosotros, sujetos cognoscentes, intentamos develarlas teorizando respecto de ese orden.

Con todo, sin poner en duda la existencia de regularidades –y, por lo tanto, de leyes- uno podría dudar de su potencia causal. Del mismo modo como la historia causal de un determinado símbolo, aun existiendo, no tiene relevancia en las propiedades causales del mismo; las regularidades o relaciones nomológicas en que entre un símbolo o estado del mundo con otros símbolos o estados del mundo nada aporta a los poderes causales del primero. El modo como están ordenados los fenómenos sólo tiene eficacia causal en la medida que existen sujetos cognoscentes capaces de reconocer dicho orden y predecir la ocurrencia de ciertos hechos a partir de otros. No obstante, la ocurrencia de cualquier fenómeno en el mundo está determinada por sus causas inmediatas y no por sus relaciones con otros fenómenos desfasados en el tiempo y el espacio. Nada importa que cada vez que haya humo tenga que haber fuego para efectos de los poderes causales contrafácticos del humo. Éstos estarán completamente determinados por sus propiedades metafísicas y no por su historia causal o por su participación en ciertas regularidades del mundo.

A modo de ejemplo, volvamos al movimiento de la cola de mi perro. Si es que esto expresa jovialidad y esto supone un efecto en el mundo sin mediar capacidad epistémica alguna, entonces debe ser debido a que su significado es una propiedad metafísica. Sin embargo, si es que la existencia de una regularidad que relaciona el movimiento de la cola con estados psicológicos de jovialidad en mi perro participa en el entramado causal del mundo esto no se debe tanto a la existencia de dicha regularidad como a las propiedades metafísicas de los fenómenos que de ella participan. Pero es difícil imaginar que el solo movimiento de cola de mi perro (viz. independiente de sus estados psicológicos) tenga entre sus propiedades metafísicas la de expresar jovialidad, por la misma razón que es difícil pensar que el movimiento vertical de la cabeza humana contenga entre sus propiedades metafísicas la de expresar aprobación. Consecuentemente, es difícil pensar que el significado es una propiedad metafísica y al mismo tiempo que es un asunto de regularidades legaliformes del mundo. Por el contrario, la relevancia en el devenir metafísico del mundo (en el cual la intencionalidad juega un rol), reitero, es un asunto local; esto es, de los fenómenos en sí, y no de sus lejanas relaciones con otros fenómenos.

Más generalmente, la similitud de la semántica informacional con (V2) arrastra también muchos de los inconvenientes que ésta enfrentaba. Por ejemplo, en una semántica informacional resulta difícil imaginar que el movimiento de la cola de mi perro posee significado de manera esencial, cuestión bastante intuitiva para la intencionalidad de las entidades mentales. Por el contrario, pareciera ser que el movimiento de la cola de mi perro puede existir sin que exprese jovialidad (bastaría que yo le moviera su cola con mi mano) a diferencia de mi concepto PERRO el cual no puede existir a menos que exprese la propiedad de ser perro.

Como sea, presumiblemente este problema radica en el mismo hecho de más arriba; esto es, que es difícil imaginar que el movimiento de la cola de mi perro posee algunas propiedades metafísicas extrañas asociadas a su significado. De ser así, expresaría jovialidad independientemente de que hallan mentes alrededor, cuestión intuitivamente inverosímil. Sin embargo, no es tan difícil imaginar que, por ejemplo, los conceptos o entidades mentales, sí poseen propiedades metafísicas extrañas de las cuales supervienen las propiedades semánticas. Y, por la misma razón, no es difícil imaginar que los conceptos poseen intencionalidad de manera esencial; esto es, dejan de significar, dejan de ser el concepto que son.

Y por si esto no bastará para mostrar que el significado no puede ser información porque la información aun existiendo no tiene relevancia metafísica, hay otro argumento independiente que nos puede llevar a la misma conclusión. En efecto, la tesis de que la información es parte de la estructura ontológica y, por lo tanto, está presente (aun cuando no constituye una propiedad metafísica) independientemente de si nos damos cuenta de ello, supone que la información es sólo aprehensible epistémicamente. Al establecer teorías sobre el orden del mundo estamos también realizando hipótesis respecto de la información presente en él. Así, por ejemplo, si existe o no una conexión nomológica entre el movimiento de la cola de mi perro y sus estados psicológicos es una cuestión metafísica del mismo modo que lo es el hecho de si el movimiento de su cola porta o no información sobre sus estados psicológicos. Lo que no es un asunto metafísico es lo que yo sé o creo saber respecto de la conexión nomológica entre esos dos fenómenos. Mi acceso a las leyes del mundo, entendidas como entidades metafísicas, es siempre mediado por mis capacidades epistémicas. Del mismo modo, mi acceso a la información metafísicamente contenida en el orden del mundo es también hipotético.

Pero si significado es información, como la entiende Dretske, entonces, reitero, sería también una cuestión presente en la estructura ontológica del mundo al cual sólo tendríamos acceso epistémico. Y, el acceso epistémico supone error. Éste no es tan grave para el caso de símbolos físicos (viz. no-mentales), pues resulta perfectamente posible que alguien no sepa o se equivoque respecto del significado de, por ejemplo, el movimiento de la cola de mi perro. Pero sí es muy grave para el caso de los símbolos mentales, pues no es aceptable es que alguien no sepa el significado de sus propios conceptos.

Puesto de otro modo, la semántica informacional establece un corte demasiado severo entre el significado de un símbolo y lo que un símbolo significa *para alguien*. El primero es un asunto ontológico, mientras que el segundo es un asunto epistémico. Yo creo que el movimiento de la cola de mi perro expresa jovialidad, pero puedo estar equivocado sobre esto. Puede que no haya ninguna conexión nomológica entre el movimiento de cola de mi perro y sus estados psicológicos. Del mismo modo, yo creo que mi concepto CABALLO *es acerca de* caballos, pero, según esta teoría, yo podría estar nuevamente equivocado. Podría ser que no existiera una conexión nomológica entre mi concepto y los caballos. No puedo tener certeza respecto del significado de mi concepto, lo cual es a todas luces inaceptable. Que el significado de mis conceptos no depende de mis capacidades epistémicas es una cosa. Que no tenga acceso directo (i.e. sin error) al significado de mis conceptos es otra completamente distinta.

El problema de fondo es el mismo que insinuábamos más arriba: la semántica informacional no distingue entre el significado de los símbolos mentales y el significado de otro tipo de símbolos. Esta teoría establece un criterio común a cualquier cosa que porte alguna forma de significado. Para mi, en cambio, la diferencia entre el significado que puede tener la palabra escrita “perro” y mi concepto PERRO es muy clara. En el primer caso, el significado de la palabra escrita “perro” no es esencial a la palabra misma. En el segundo, el significado de mi concepto PERRO es esencial a él. Así, si, por ejemplo, hubiese una catástrofe nuclear y desaparecieran todas las formas de vida en el planeta, el objeto físico asociado a la palabra “perro” no tendría significado alguno. Dicho de otro modo, una palabra escrita puede sobrevivir sin que lo haga su significado. Por el contrario, un concepto PERRO no necesita ser interpretado. Si no significa perro no puede sobrevivir como tal.

La conclusión que se sigue de lo anterior es que el significado de un símbolo físico es de naturaleza distinta del significado de un símbolo mental. Mientras que es inaceptable que una teoría del significado conceptual lo conciba como dependiendo de las capacidades epistémicas de los sujetos, esto es perfectamente tolerable para el significado de los símbolos físicos. La Restricción Epistémica sólo se aplica, por lo tanto, a la semántica de las representaciones mentales y no a la semántica de otro tipo de símbolos (e.g. físicos). En consecuencia, la semántica informacional debe, a mi juicio, aceptar una noción de información que, aunque mediada por capacidades epistémicas, al menos posee relevancia. De este modo, deberá conformarse con explicar el significado de los símbolos físicos (i.e. no-metales). Una teoría como esta, restringido su alcance, me parece no sólo sólida desde la perspectiva teórica sino que incluso atractiva –probablemente de allí provenga su reciente popularidad. Creo, de hecho, que ésta es la mejor teoría de que disponemos para explicar la semántica de los símbolos físicos.

Lamentablemente, sin embargo, arreglar la semántica informacional nos deja sin una teoría de la intencionalidad –en el sentido en que Brentano usaba el término. Habrá, pues, que construir otra teoría que explique el significado de los conceptos y que explique también por qué éste posee una naturaleza tan especial y distinta del significado de otros símbolos. Intentar desarrollar una teoría tal será otro motivo para hacer el intento que haremos en los capítulos restantes.

## **Biosemántica**

En el transcurso de los capítulos 4 y 5 estudiaremos con detenimiento las semánticas del rol funcional, las cuales pretenden individuar funcionalmente el contenido conceptual. El concepto mismo sería, a su vez, individuado por su contenido, hecho compatible con la intuición de que los conceptos poseen su contenido esencialmente. Sin embargo, como veremos, este tipo de semánticas adolece una serie de inconvenientes teóricos, siendo el más grave el holismo semántico y, de allí, su compromiso con el solipsismo.

Podría decirse que la biosemántica (Millikan 1984) pertenece al conjunto de las semánticas del rol funcional sólo en virtud de que define las propiedades semánticas en términos de la noción de función. Sin embargo, ésta rara vez es clasificada en este grupo puesto que la noción de función a la que apela es radicalmente distinta de la noción matemática tan características de las semánticas referidas de este modo.

En cambio, como lo indica su nombre, la biosemántica intenta establecer una noción de función heredada de aquélla usada en biología. La idea, observa Sober (1985), es que una noción de función teleológica, como la ocupada en las ciencias biológicas, ahorraría a una teoría funcionalista de la mente un sinfín de inconvenientes (e.g. el problema de la Población de China o el del Homunculismo; ver Capítulo 4). Del mismo modo, razona Millikan, podríamos observar los frutos de adscribir a dicha noción en una teoría del contenido. El único problema que debe enfrentar es que no existe una formalización científica o filosófica suficientemente rigurosa para definir la noción teleológica de función, hasta el momento entendida de modo estrictamente intuitivo.

En consecuencia, el trabajo realizado por Millikan tiene una doble dimensión. Por un lado, se trata de un intento inaudito hasta no hace mucho de formalizar una noción teleológica de función, con la cual podrían resolverse un montón de dificultades fundacionales de las teorías funcionalistas de la mente. Por otro lado, someter dicha noción a una prueba de fuego como es el problema del contenido intencional. La idea es individuar el contenido funcionalmente y, por medio de alterar la noción de función, rehuir ciertos problemas atribuibles a las versiones clásicas de las semánticas del rol funcional (Fodor y Lepore 1992).

Grosso modo, Millikan intenta definir la noción de función apelando a otra de normalidad. De este modo, la función de un órgano en un ente biológico estaría dada por el desempeño del organismo en una situación normal, acorde con el proceso evolutivo que terminó con el

desarrollo de su especie. La función del corazón sería, pues, bombear sangre al cuerpo, gracias a que esto es para lo que el corazón ha sido moldeado (supuesto) a través de la evolución. Esto a pesar de que en ocasiones éste pueda realizar erróneamente su función o simplemente no realizarla, debido a que, dadas las circunstancias ambientales *normales*, la evolución ha determinado que desempeñándose de esa manera el corazón permitirá la adaptación del individuo a dichas circunstancias. Consecuentemente, cuál es la función de un cierto aparato no sería una cuestión de hecho, sino que dependería de la historia del aparato y, presumiblemente, de la de toda su stirpe. Millikan ha llamado “función propia” a esta noción independiente del estado de cosas factual, para distinguirlo de la noción habitual de función.

Al aplicar esta idea al problema de la intencionalidad el resultado es previsible. El significado, entendido como mapeo desde estados mentales a porciones de mundo, sería un asunto de la función propia de los estados mentales. En otras palabras, la función propia de una creencia sería, bajo esta concepción, referir a porciones de mundo. La función propia de un concepto, por su parte, referir a un objeto. Traducido a lenguaje coloquial, por medio de la evolución se han desarrollado ciertas entidades mentales al interior de los individuos llamadas creencias cuya función es corresponder a una porción de realidad. Así de simple.

La principal virtud que los partidarios de este tipo de explicación ven en ella es que ésta conseguiría mostrar la intencionalidad bajo preceptos comunes con los de la biología. Así, las creencias falsas serían equiparables a órganos o dispositivos biológicos que funcionan mal. En otras palabras, una creencia falsa sería una que no desempeña su función propia, pero que, sin embargo, posee significado toda vez que, desempeñela o no, posee una función propia. Con todo, esta reducción del rol de las creencias al lenguaje de la biología sustenta la apariencia de la intencionalidad como fenómeno natural.

Del mismo modo, podríamos ver la misma explicación desde la perspectiva de los mecanismos de fijación de creencias antes que desde las creencias mismas. Esto es, la función propia de estos mecanismos sería producir creencias verdaderas, siendo el caso que fallan en desempeñar su función propia cuando producen creencias falsas. Aun así se requeriría una descripción alternativa de dicha función, bajo pena de explicar las propiedades semánticas recurriendo a otra típicamente semántica, como es la verdad. Ésta descripción provendría del desempeño evidencial de dichos mecanismos en *condiciones normales* (u *óptimas*). Así, la función propia de los mecanismos de fijación de creencias sería producir lo que producen bajo dichas condiciones (i.e. creencias verdaderas).

Analizando la capacidad de este tipo de teorías teleológicas del contenido para tratar el problema ya clásico de la disyunción, Fodor (1987) ha dado con algunas dificultades que éstas enfrentarían. Según él, no es claro que exista la posibilidad de especificar unívocamente cuáles serían las *condiciones normales* en las que los mecanismos de fijación de creencias desempeñarían su función propia. Por el contrario, pareciera que las condiciones en las que los mecanismos producen creencias verdaderas son distintas cuando éstas involucran, por ejemplo, objetos pequeños de las que se requerirían cuando involucran objetos grandes. En el primer caso, presumiblemente, las condiciones adecuadas serían, entre otras, cercanía del sujeto cognoscente a los objetos; mientras que en el segundo caso, se requeriría, *ceteris paribus*, lejanía del mismo a los objetos. Esto sugiere la desagradable idea de que no es posible especificar las *condiciones normales* sin que medie una descripción intencional de la situación.

Sin embargo, este argumento no es demostrativo. Más aun, lo que los partidarios de las semánticas teleológicas supuestamente tienen en mente es la posibilidad –en principio al menos– de especificar las *condiciones normales* en términos evolutivos, comunes a cualesquiera creencias en particular y, *a fortiori*, no-intencionales. Claro está, ninguno de ellos ha entregado esa especificación, lo cual, lamentablemente, resta credibilidad a la teoría, aunque no consigue derribarla teóricamente.

Personalmente, creo que esta teoría podría ser muy atractiva como teoría de los mecanismos de fijación de creencias, pero no es igualmente buena como teoría del contenido. La razón, es que no responde lo que se supone debería responder si fuera de este último tipo; a saber, de dónde superviene la propiedad de los estados mentales de *ser acerca de* porciones de realidad. Si esto es así, la teoría no satisface ninguna de las restricciones anteriormente mencionadas (ni la Semántica ni la Epistémica), pues no satisface ninguno de los antecedentes: no explica las propiedades semánticas.

En efecto, al decir que la función propia de una creencia es corresponder con porciones de mundo esta teoría sólo está constatando el hecho que se desea explicar; a saber, que los estados mentales son comúnmente *acerca del* mundo. Pero la pregunta sigue en pie: ¿Cuál es el *grounding* metafísico de esta rara propiedad de los estados mentales de corresponder comúnmente con porciones de mundo? O, en otras palabras, ¿de qué propiedades metafísicas proviene la intencionalidad? Suponer una teoría de la correspondencia entre estados mentales y porciones de mundo no es explicarla.

Proponer que la propiedad de –digamos- las creencias de *ser acerca de* porciones de mundo equivale a su función propia sin aportar una explicación independiente de cuál es esta función es simplemente introducir un neologismo para un fenómeno que ya tenía nombre (i.e. “intencionalidad”). No hay ningún aporte en eso para una teoría del contenido. Esto es precisamente lo contrario de lo que ocurre en otros casos análogos en la Biología. El mismo procedimiento consigue aportar toda vez que allí normalmente se especifica la función un cierto dispositivo biológico de manera independiente. La función propia de corazón, por ejemplo, sería bombear sangre –sentencia que no contiene implícitamente referencia a corazones.

En resumen, la biosemántica resulta ser una excelente teoría excepto en lo de semántica. Claramente si no reúne este requisito ni siquiera habremos de preocuparnos de ella como posible explicación competente de la intencionalidad. En lo que resta comenzaremos con nuestro propio intento de entregar una explicación metafísica y naturalista de la intencionalidad desde una perspectiva desconocida hasta la fecha. Dejaremos en manos del lector la tarea de poner el resultado en la balanza junto con otras ofertas alternativas para ver cuál pesa más.

## Capítulo 3

Hume observó que los argumentos de Berkeley en favor de su Escepticismo solipsista no admiten la menor réplica y no producen la menor convicción. Algo parecido podría decirse de la postura simétricamente inversa: el Conductismo Radical. Descartar la actividad mental de plano nunca contará con el beneplácito de nuestra íntima conciencia –si se me permite la expresión. Como sea, la categoría ontológica de lo mental es, sin duda, uno de los problemas más enigmáticos de la historia de la filosofía. Durante el último siglo, sin embargo, ha habido avances sustantivos en, al menos, el esclarecimiento de la naturaleza del problema. Los filósofos de la mente han determinado, con mayor o menor adhesión, que la base de aquello llamado mental radica en una íntima relación entre sujeto y mundo, la cual, a su vez, posee un rasgo distintivo conocido como intencionalidad (ver capítulo 1). Hablar sobre la categoría ontológica de lo mental es, en último término, equivalente a hablar sobre la categoría ontológica de la intencionalidad.

La existencia o inexistencia de la intencionalidad como fenómeno en sí mismo ha sido motivo de álgida disputa en la Filosofía de la Mente<sup>8</sup>. Este hecho, sin embargo, no resulta extraño si se revisa el conjunto de las razones que soportan una y otra postura. Si, por un lado, existen supuestos ontológicos que nos parecen incuestionables y que hacen inaceptable la existencia independiente de los fenómenos intencionales; por otro, somos irrenunciables testigos de nuestra actividad intencional: negar su existencia no conseguirá que esta íntima convicción desaparezca. En segundo lugar, los intentos por hacer compatibles los fenómenos intencionales con aquellos supuestos ontológicos no han rendido frutos. Dichos fenómenos parecieran ser de una naturaleza incompatible con las concepciones derivadas del materialismo. Esta aparente ininteligibilidad del problema ha sido el mejor argumento para quienes han defendido, concientes o no de él, posturas dualistas desde hace siglos. Intentar naturalizar la Intencionalidad es, pues, una apología del materialismo.

---

<sup>8</sup> Ver, por ejemplo, Borst, C.V. 1983

## Materia e Intencionalidad

Hacer una teoría de naturalización de la intencionalidad es establecer una conexión explicativa entre dicha noción y los supuestos sobre los cuales se sientan las Ciencias Naturales. Para tener alguna esperanza de alcanzar este objetivo lo que no debemos hacer es contrariar aquellos supuestos.

Que el Materialismo Ontológico es un supuesto generalmente tácito aunque presente en toda Ciencia Natural (acaso en toda forma de actividad científica) es un hecho difícilmente cuestionable. La Filosofía de las Ciencias Naturales avala este hecho al plantear la teoría de los tres niveles explicativos. Según ésta, la Biología, la Química y la Física, constituyentes de las Ciencias Naturales, estarían, en ese orden, en una relación de dependencia explicativa. Así, los axiomas de la Biología (única donde ha habido quienes han planteado su irreductibilidad a la Física, aunque sin demasiado fundamento<sup>9</sup>), deben poder explicarse en términos de quienes la suceden en el escalafón, a saber, la Química y la Física. En ese sentido la primera es reducible, al menos en principio, a las otras dos. Del mismo modo, la Química debe ser reducible, en principio, a su sucesora, la Física. Mientras que esta última, al ser el nivel fundamental, deberá referir todos sus principios directamente a lo material y sus relaciones. Con todo, quizá la prueba más patente del compromiso materialista de las Ciencias Naturales es que nadie que pretenda tener alguna credibilidad dentro de la comunidad científica postularía una teoría dual del universo, donde se produjo una evolución paralela de dos sistemas constitutivos de éste; esto es, que la parte material se desarrolló de una cierta manera mientras que la otra, cualquiera que esta sea, lo hizo de modo independiente. Más allá de ser o no defensor del Materialismo –que lo soy-, la presente investigación exige, pues, la adhesión al principio de que todo lo que hay en el mundo es materia y no hay más hechos en él que interacciones entre ella<sup>10</sup>.

Por otro lado, si el hombre es, como las Ciencias Naturales suponen, parte y producto de este mundo material, entonces, todo lo que hay en él ha de ser también material. De no aceptar

---

<sup>9</sup> Ver Chuaqui, B. 1995

<sup>10</sup> Para ser justos con la Teoría de la Relatividad deberíamos agregar a aquello de que se compone el mundo la Energía. Ésta, en todo caso, corresponde a un posible estado de la materia.

esta afirmación, habría, pues, que explicar en qué momento de la evolución del hombre o de la de sus antepasados –desde las partículas replicantes (quizá desde los aminoácidos, azúcares, purinas y pirimidinas) hasta el Hombre de Cromañón (Maynard Smith, J. & Szathmáry, E. 2001)- las interacciones materiales dieron lugar a fenómenos que, al no ser materiales, establecieron aquella súbita dimensión de naturaleza inmaterial del hombre; o, alternativamente, habría que refutar la Teoría de la Evolución de las Especies o la Biología Celular. Ninguna de estas alternativas, sin embargo, demuestra alguna plausibilidad. En definitiva, pocas dudas pueden caber que tal línea argumental está decididamente reñida con las Ciencias Naturales. Habremos de aceptar, en consecuencia, que si el hombre desciende del mundo material, todo lo que hay en él ha de ser también material, y que, por lo tanto, la intencionalidad es, en último término, alguna forma de interacciones materiales.

Y, a pesar de todo, la tentación persiste...

Somos testigos privilegiados e irrenunciables de la actividad intencional. Todos los días experimentamos la nuestra propia y por medio de ella entendemos las relaciones humanas, incluida, por supuesto, nuestra relación con los restantes individuos. La llamada Psicología Popular es la mejor teoría con que contamos para explicar y organizar el curso de los acontecimientos humanos (Fodor 1975). Cuando caminamos por la calle y alguien nos pregunta qué hora es, suponemos que ese alguien, al proferir dicha pregunta, expresa inquietud *acerca del* momento del día en que nos encontramos. Cuando en la universidad comentamos con alguien lo enigmático que nos parece un cierto problema filosófico, nuestro interlocutor supone que intentamos reflexionar *acerca de* dicho problema. Sin dichas suposiciones no tendría caso mirar el reloj e informar la hora que es o tratar de abordar aquel intrigante problema filosófico. Sin atribuir intencionalidad a los pensamientos ajenos y propios el mundo de las relaciones humanas en que se enmarca nuestra vida simplemente no funcionaría (o, al menos, no como ha funcionado hasta ahora). ¿Cómo podremos entonces negar la existencia de la intencionalidad?

Nada de esto sería problemático si pudiéramos explicar el fenómeno de la intencionalidad con las mismas herramientas teóricas con que explicamos los fenómenos físicos. Esto, sin embargo, hasta la fecha ha sido imposible; las dificultades que aparecen al intentarlo son múltiples (Ver capítulo 2). Consideremos, a modo de ejemplo, el sencillo caso en que alguien, digamos, Juan, nos dice que Pedro, un amigo común, está en su casa. Hay un conjunto de problemas relacionados con la enunciación por parte de Juan de la oración ‘Pedro está en su

casa'. En primer lugar está el problema del locus de la intencionalidad. Hay quienes han planteado que los lenguajes naturales son portador primitivos de intencionalidad. Esto equivale a decir que, por ejemplo, el símbolo 'Pedro', ya sea sonoro, escrito, u otro, posee intrínsecamente la capacidad de referir a Pedro. Creo que quienes han planteado tal alternativa han pasado por alto la posibilidad nomológica (para que decir metafísica) de que los padres de Pedro hubieran decidido bautizarlo como Pablo, en vez de Pedro. En ese caso, dirían los mismos, el símbolo 'Pablo' referiría intrínsecamente a Pedro, mientras que el símbolo 'Pedro' no referiría a Pedro. Pero entonces, ¿puede haber algo intrínseco al símbolo 'Pedro' que refiere a Pedro? No, pues si así fuera, ¿cómo es que tal capacidad depende de un hecho contingente (viz. la decisión de llamar 'Pedro' a Pedro)? El lenguaje es una convención hecha para poder expresar nuestros pensamientos a la cual nos sumamos en la medida que pertenecemos a una comunidad. Éstos son los portadores primitivos de Intencionalidad, el lenguaje un mero vehículo de contenido intencional (Fodor 1998). Para quienes viven en España, el Español es útil para expresar sus pensamientos; para quienes viven en Inglaterra, el Inglés. Esta es la razón por la cual los españoles hablan Español y no Inglés, mientras que los ingleses hablan Inglés y no Español.

Siendo así las cosas, olvidemos la enunciación por parte de Juan de la oración 'Pedro está en su casa' y concentrémonos en el pensamiento de Juan sobre el hecho de que Pedro está en su casa. Según cree la gran mayoría de quienes participan de la Ciencia Cognitiva, dicho pensamiento posee una estructura (viz. una sintaxis), la cual es análoga a la estructura sintáctica en términos de Sujeto y Predicado que usamos para analizar el lenguaje. De este modo, un pensamiento tal posee como uno de sus componentes la representación mental PEDRO<sup>11</sup>. Ésta, según han planteado algunos identificados con las llamadas Arquitecturas Conexionistas, correspondería a una cierta configuración neuronal en el cerebro, y, al reducirla a un fenómeno biológico (y, por lo tanto, físico), es por medio de esta última como debería ser estudiada (Ver, por ejemplo, Churchland P.M. 1981).

De lo anterior estoy dispuesto a aceptar la primera sentencia, pero no la segunda. La razón es que, más allá de si dicha representación mental corresponde o no a una cierta configuración

---

<sup>11</sup> Adoptaremos la convención estipulada en *Concepts: Where Cognitive Science were wrong*, de Fodor, para referirnos a la representación mental (PEDRO), a la expresión del lenguaje ('Pedro') y al objeto referido (Pedro).

neuronal, ese hecho no resulta muy iluminador respecto de por qué la representación mental PEDRO está dirigida a Pedro. El problema de entender por qué dicha representación mental es *acerca de* dicho individuo sería reemplazado por el problema de entender por qué dicho fenómeno biológico, que es la configuración neuronal asociada, es *acerca de* Pedro. Esto, aun cuando podría significar un gran avance hacia naturalizar los objetos (e.g. representaciones) mentales, no representaría ningún avance en relación con el problema de naturalizar la intencionalidad que dichos objetos podrían poseer. He aquí el verdadero problema.

Olvidémonos, en consecuencia, de la pregunta por si la representación mental PEDRO corresponde o no a una cierta configuración neuronal. Las representaciones mentales son, en todo caso, portadoras de contenido intencional, lo que equivale a decir que tienen la capacidad de ser *acerca de* objetos o propiedades del mundo. Esta capacidad, según las llamadas Teorías Representacionales de la Mente (TRM), es intrínseca al hecho de ser representación mental. Esto es lo que quiere decir Fodor en la segunda tesis definitoria de las Teorías Representacionales de la Mente, al plantear que las representaciones mentales son las portadoras primitivas de contenido intencional (Fodor 1998; p.7). En otras palabras, se intenta establecer que, ontológica y explicativamente, la intencionalidad está ubicada en las representaciones mentales mismas antes que en las actitudes proposicionales (viz. deseos o creencias) en que éstas aparecen involucradas. Las últimas heredarían su intencionalidad de las primeras. Esta distinción, a diferencia de la de los conexionistas, sí es determinante para el problema de la naturalización de la intencionalidad.

Habría que preguntarse, entonces, qué hay en las representaciones mentales que permite a éstas ser *acerca de* algo. Distintas respuestas se han dado al interior de las Teorías Representacionales de la Mente a esta pregunta (ver capítulo 2). Ninguna de ellas, sin embargo, ha conseguido hacerlo convincentemente en términos no-intencionales y no-epistémicos y al mismo tiempo compatibles con las Ciencias Naturales. La razón, según creo, es que al considerar a las representaciones mentales como las portadoras primitivas de la intencionalidad se ha establecido el prejuicio de que una explicación de esta propiedad debe realizarse con independencia de la participación las representaciones mentales en las actitudes y, en general, en la estructura interna de los organismos dotados de mente. En consecuencia, todas las hipótesis que puedan hacerse sobre la naturaleza de la relación semántica estarán demasiado lejos de lo que puedan haber dicho las Ciencias Naturales sobre el hombre y el mundo. Las semánticas causales o informacionales discutidas en el capítulo 2 son ejemplos de teorías que buscan una explicación de las propiedades semánticas en relaciones directas entre

el símbolo mental y el objeto (o propiedad) representado. La estructura del organismo portador del símbolo mental no juega ningún rol en dichas explicaciones.

En resumen, después de dar una vuelta por algunos problemas asociados con la naturalización de la intencionalidad, vemos por qué resulta ser tan tentador acudir a un dualismo ontológico. Sin embargo, no debemos morder la manzana. En los capítulos que siguen intentaremos una salida a tan intrincado problema. La novedad del intento radicará en el hecho que, según creo, ha imposibilitado resolver el enigma: la suposición de que la intencionalidad debe ser explicada con independencia de la estructura del organismo portador de actividad mental. Postularemos precisamente lo contrario, a saber, que las representaciones mentales reciben su intencionalidad de las actitudes proposicionales y de su rol en la conducta, entendida como proceso biológico. Y, en la medida que esta explicación va más lejos que las meras representaciones mentales, intentaremos su naturalización, completando así la secuencia.

## **Los Niveles Explicativos**

Quién sabe por qué el hombre tiene la costumbre de clasificar. Desde que aparece el lenguaje en la vida de los hombres su necesidad de clasificar objetos, problemas, situaciones o lo que sea no se ha detenido jamás. Acaso sea parte de una conducta adaptativa, acaso sea inherente a los animales cerebrados. Como sea, existe una rara conexión entre clasificar y hacer conocimiento. Más aun, pareciera que hacer conocimiento es, al menos en parte, ordenar y clasificar los objetos del mundo de alguna forma. Esta es seguramente la razón por la que el conocimiento se ha construido por niveles. En la era moderna esto es especialmente claro. Existen las Ciencias Naturales, las Ciencias Sociales, las Ciencias Computacionales, la Zoología, las Ciencias del Hombre, etc., todas ellas con un sinfín de clasificaciones internas. Las Ciencias Naturales se dividen en Física, Química y Biología; las Sociales en Psicología, Sociología, Antropología; y todas ellas se subdividen, a su vez, en más y más ramas como la Física de Sólidos, de Partículas, la Química Orgánica y la Inorgánica, la Biología Celular y la Molecular, la Genética, etc. Estas clasificaciones responden, en todo caso, a distintos requerimientos. Algunas de estas divisiones tienen que ver con definiciones respecto de los objetos de estudio, otras con los supuestos sobre los que se fundan y muchas respecto del ámbito relevante de análisis. El conocimiento se construye, así, por parcelas, y, no obstante, la convicción de que el objeto último de estudio, que es el mundo, es uno solo, ha hecho soñar a los hombres, y especialmente a los epistemólogos, que algún día contaremos con una teoría

unificada del mundo. Soñar es gratis, seguimos estando muy lejos de una teoría tal, acaso ésta sea sólo una quimera. Sin embargo, este tipo de convicciones tiene un efecto práctico en la forma de hacer conocimiento, a saber, el valor agregado que tienen aquellas teorías que logran situarse en armonía con el resto de las teorías circundantes (Laudan 1981).

Por esta razón construir una teoría de la intencionalidad naturalizada es importante. Por esta razón también, y por lo lejano que parecen los fenómenos intencionales del ámbito de estudios de las Ciencias Naturales, construirla por niveles talvez sea lo más saludable. El método de construcción que proponemos en la presente Tesis consta, en consecuencia, de dos pasos. En primer lugar, intentaremos explicar la conducta teleológica propia de los seres vivos en términos de una concepción mecanicista (en consecuencia, materialista) de éstos. Para esto haremos uso de la Teoría de la Autopoiesis de Humberto Maturana y Francisco Varela, la cual tiene la notable capacidad de mostrar el comportamiento teleológico como un efecto producido en el observador tras el cual subyace una realidad mecánica (Maturana y Varela 1964). Así, como se verá en detalle en el capítulo 4 del presente trabajo, los sistemas biológicos no serían más que sistemas mecánicos moldeados a través de la evolución de modo tal que su funcionamiento mantiene su propia organización interna constante. De esta manera, todo comportamiento de un sistema tal que parezca dirigido a la obtención de un objetivo (viz. teleológico) no es más que la expresión del funcionamiento interno de una máquina que al cambiar las condiciones del entorno realiza cambios compensatorios manteniendo su organización estable. Luego, a partir de esta notable idea formularemos una caracterización de la conducta intencional que permitirá entenderla como uno de los procesos del fenómeno de la vida de los organismos dotados de cerebro. Esto es lo que llamaremos la psicopoiesis de dichos organismos.

A continuación, en el capítulo 5, intentaremos explicar la intencionalidad en términos de la noción de psicopoiesis antes desarrollada. Esto permitirá entender la intencionalidad como exclusiva de los organismos vivos dotados de mente. Adicionalmente, la intencionalidad no será concebida como fenómeno localizado en una parte específica de la composición del hombre, sino como parte del proceso completo que lo define como ser vivo. En particular, a pesar de tener una estrecha relación con el cerebro (viz, sistema nervioso central), postularemos que la intencionalidad no es aprehensible por medio del análisis del funcionamiento interno del cerebro, sino que, más bien, es expresión de una característica más profunda aun de los seres vivos: su autopoiesis; sumada ésta a la función del cerebro en la conducta (viz. psicopoiesis) de los individuos que exhiben dicha característica. De modo más

preciso, diremos que la intencionalidad es una expresión del comportamiento psicopoiético de ciertos seres vivos que han alcanzado un mayor grado de interacción con el medio circundante, a saber, aquellos que poseen sistema nervioso central. Específicamente, postularemos que las actitudes proposicionales son expresión directa del comportamiento autopoiético de los seres cerebrados. Consecuentemente, las razones que explican la intencionalidad en los individuos que muestran comportamiento teleológico serán aquellas que expliquen el cerebro como órgano dinamizador de dicho comportamiento en términos de sus interacciones con el entorno. Todas ellas de carácter filogenético.

La resultante de esta doble secuencia explicativa a exponerse en el capítulo 6 puede no ser muy sencilla de asir, pero ¿qué puede esperarse de la explicación de uno de los fenómenos más complejos que experimenta uno de los organismos más complejos que nuestro universo material ha producido? Cualquier teoría que dé cuenta de la actividad mental en términos de una concepción material del mundo no ha de ser juego de niños. Sin embargo, acaso esta visión global del asunto no es lo único que debe considerarse al momento de juzgar un intento como el que se está llevando a cabo. Más aun, podríamos esperar que en algunas décadas aparezcan otras teorías sobre la materia que explicarán el fenómeno de modo más preciso y quizá con mayor simpleza, pero esto no aminora la importancia del actual intento respecto de establecer la posibilidad en principio de explicar el ámbito de lo mental en términos del material. En efecto, si los dos pasos explicativos propuestos anteriormente no presentan ningún error de coherencia lógica, tendremos así una teoría, buena o mala, capaz de producir una explicación razonable de lo mental en términos de lo material y, con ello, se habrá establecido la posibilidad de realizar tal empresa, descartando, a su vez, el mejor argumento sobre el que se han sostenido las posturas dualistas. Más allá de su efectividad para representar la realidad, la teoría tiene valor como apología del Materialismo.

## **Intencionalidad, Neurobiología y Reduccionismo**

La Intencionalidad ha jugado un papel central en la discusión al interior de la Ciencia Cognitiva sobre el modo como debería enfrentarse el estudio de la Cognición. Los partidarios de las llamadas Arquitecturas Conexionistas han defendido no sólo posturas reduccionistas respecto de los estudios realizados por la Psicología y Filosofía de la Mente al ámbito de la Neurobiología, sino que ha habido quienes han llegado más lejos al plantear la eventual desaparición de dicha forma de encarar el problema en favor de esta última. Sin embargo, a

pesar de que diversos argumentos han sido esgrimidos para apoyar dichas tesis, la indomabilidad teórica propia de la noción de intencionalidad ha estado a la orden del día para la argumentación de estos filósofos comunes en el Sur de California. Con todo, a pesar de este tipo de sugerencias, como la hecha por Churchland P. y Sejnowsky (1990) sobre la posibilidad de reducir la intencionalidad (*Aboutness*) al lenguaje de la Neurobiología, la presente teoría de la Naturalización excluye explícitamente esta posibilidad y, en general, toda iniciativa tendiente a reducir la Psicología Cognitiva a la Neurobiología –ni hablar de las propuestas eliminativistas.

Crear en la posibilidad de la naturalizar la Intencionalidad exige, en efecto, creer en la posibilidad de reducir la Psicología a las Ciencias Naturales. Pero Ciencias Naturales no es lo mismo que Neurobiología. Si hay una teoría a la cual puede reducirse la intencionalidad esta es la Teoría de la Autopoiesis de Maturana y Varela; y, obviamente, no a teorías sobre el funcionamiento interno del cerebro. La razón, en lenguaje computacional, es que la capacidad de representar de la mente humana no tiene que ver con las propiedades de la máquina sobre la cual se implementan los procesos cognitivos, sino que con una propiedad enraizada mucho más profundo en la estructura propia del ser vivo como tal: la Autopoesis. En principio es perfectamente posible imaginar una máquina autopoietica (viz. ser vivo) teniendo actitudes proposicionales implementadas sobre un computador serial, puesto que la capacidad para tener actitudes proposicionales (i.e. de ser *acerca de*), depende sólo de la estructura autopoietica del ente y de la posesión por parte de éste de un órgano procesador de información funcional a dicha estructura y suficientemente eficaz. Esta idea es, sin dudas, identificable con las llamadas Arquitecturas Seriales de la Mente, y, aunque sea difícil de creer, especialmente coherente con algunos de los requerimientos de las Teorías Representacionales de la Mente pues permite separar el contenido intencional de los procesos cognitivos, cuestión central en dicha propuesta (Pylyshyn 1984).

Pues bien, una vez descartada la reducción de la intencionalidad a la Neurobiología queda la pregunta más general por la reducción de la intencionalidad y de la Ciencia Cognitiva al lenguaje de la Autopoiesis. La respuesta a dicha pregunta parece encontrarse en varias razones que muestran la banalidad de dicha reducción. Algunas de ellas son las siguientes:

1. La Teoría de la Autopoiesis, a diferencia de la Neurobiología, no supone el establecimiento de un ámbito de investigación científico, sino que, más bien, ésta es una teoría meramente explicativa de la vida en términos mecanicistas (en consecuencia, materialistas). Al ser una

teoría que sirve de nexo entre dominios, más que servir para reducir otras teorías a ella, constituye en sí misma la posibilidad de reducción (y en ningún caso eliminación) de los fenómenos biológicos al ámbito material. Es una teoría *cuasi*-filosófica en la medida que responde a la necesidad del hombre de dar coherencia a su conocimiento; no tiene sentido reducir nada a ella.

2. El lenguaje intencional muestra eficacia y economía para captar las generalizaciones relevantes al ámbito del comportamiento humano (Pylyshyn 1984). Que el comportamiento humano sea explicable por medio de la Autopoesis no aporta mucho más respecto de predicciones más allá de la posibilidad de excluir tesis dualistas sobre el hombre.

3. La separación del contenido intencional de los procesos cognitivos antes mencionada establece a su vez la independencia de la Ciencia Cognitiva en general de la Teoría de la Autopoesis. Que, por ejemplo, la percepción que tengo de un perro en la calle sea *acerca de* dicho perro no dice nada sobre el proceso mismo de percepción, sino que sólo da cuenta del hecho que yo, en tanto máquina autopoietica, debido a los cambios en mi entorno producidos por la presencia del perro (cuya detección constituye el proceso mismo de percepción) realizo una serie de compensaciones de mi organización interna que determinan la estabilidad de la misma. Los procesos cognitivos, en consecuencia, han de ser estudiados con bastante independencia del hecho que las representaciones involucradas en ellos sean *acerca de* algo. Si la Neurobiología tiene algo que decir aquí o si la implementación de dichos procesos puede realizarse en cualquier procesador de información, paralelo o no, es harina de otro costal.

## Capítulo 4

Uno de los hechos que más me intrigó durante los pasados años es la aparente incompatibilidad del comportamiento teleológico que exhiben los seres vivos con una concepción ontológica materialista (y, a la vez, mecanicista) del mundo. La supuesta contradicción se vislumbra al considerar la relación entre las formulaciones de estas nociones. El postulado del materialismo ontológico puede expresarse como sigue:

A1) No hay nada en el mundo que no sea materia.

A2) Los fenómenos del mundo son en último término interacciones entre materia.

A3) Dichas interacciones pudiendo ser múltiples (gravitacionales, eléctricas, etc.) se rigen por la ley Causa/Efecto.

La ley Causa/Efecto, a su vez, propone que todo efecto queda completamente determinado por su(s) causa(s); lo cual supone una concepción particular respecto de la forma como se determinan los fenómenos del mundo. Esta concepción puede, pues, retratarse en dos postulados fundamentales sobre el orden de determinación temporal y espacial de los fenómenos:

A3(a) (*Determinación Temporal*) Las causas anteceden temporalmente a los efectos.

A3(b) (*Determinación Espacial*) Los efectos no están contenidos espacialmente en las causas.

La primera consecuencia de la ley Causa/Efecto, que hemos llamado la *Determinación Temporal*, es de conocimiento común, pues fue observada por Hume y no amerita mayor comentario. La segunda, que hemos llamado *Determinación Espacial*, no ha sido considerada habitualmente en la literatura relacionada. La idea que intento expresar con ella es, sin embargo, más que sencilla y aceptable. Cuando se tiene un sistema material complejo con una cierta dinámica, entonces el comportamiento de dicho sistema está determinado por el comportamiento de sus componentes y no a la inversa. Así, por ejemplo, cuando observamos un derrumbe en una ladera de un cerro, el curso de éste está determinado por las interacciones

causales de las piedras que lo componen y no es el curso que sigue el derrumbe lo que determina el curso de acción de las piedras.

Esto no quiere decir que siempre las causas actúen en un ámbito espacial más reducido que los efectos, pues esto depende de cuales sean los efectos relevantes para el observador. Por el contrario, hay un sinnúmero de ejemplos en que ni las causas están contenidas en los efectos ni los efectos en las causas. Uno muy simple podría ser el caso del movimiento de una bola de billar, el cual es causado por el choque de otra bola en movimiento que en ningún caso está incluida en la primera (Salvo, por supuesto, que uno quiera considerar la fenomenología del conjunto de bolas). Sin embargo, cuando se tiene un sistema de objetos compuesto de otros objetos, o cuando se quiere considerar así, las interacciones causales entre sus componentes son las que determinan los fenómenos asociados sistema y no al revés.

En segundo lugar, tenemos el postulado del comportamiento teleológico de los seres vivos el cual puede expresarse como sigue:

B) El rasgo característico de los organismos vivos es que actúan en pos de la consecución de propósitos (*goal-directed behavior*)

Actuar en pos de la consecución de un propósito supone, sin embargo, una forma de determinación de los fenómenos completamente distinta a la anterior, toda vez que lo que determina un cierto comportamiento (viz. el propósito) es temporalmente posterior al mismo.

Pues bien, si los seres vivos somos parte del mundo –cuestión que me parece indiscutible- y si el mundo se rige según los enunciados A1-A3 entonces, los hechos que determinan la ocurrencia de un cierto comportamiento en un ser vivo (e.g. una acción) deben ser temporalmente anteriores a éste, en virtud de A3(a). Por otro lado, si aceptamos que los seres vivos exhiben comportamiento teleológico –cuestión que me parece igualmente indiscutible- entonces los hechos que determinan la ocurrencia de un cierto comportamiento deben ser temporalmente posteriores a éste, en virtud de (B). Pero los hechos que determinan un cierto comportamiento no pueden ser anteriores y a la vez posteriores a éste, de donde se sigue que ambas concepciones, una sobre el mundo y la otra sobre los seres vivos, no pueden sostenerse simultáneamente. En otras palabras, la noción de teleología invierte el orden temporal de la determinación de los fenómenos que una concepción materialista requiere.

Otra noción considerada usualmente como teleológica es la de función. Puede distinguirse, sin embargo, dos concepciones de función una teleológica y otra matemática (Sober 1985). La primera es la que usamos cuando cotidianamente decimos que la función del corazón es bombear sangre al cuerpo. La segunda, mientras tanto, es la que usamos cuando identificamos un objeto en términos de ciertos poderes causales en una estructura de estados que finalmente produce una cierta salida (output). Más allá de esta distinción, ambas comparten una característica fundamental, a saber, violan A3(b). En otras palabras, en ambos casos se determina un cierto objeto acudiendo a fenómenos del sistema que lo contiene, obviando el orden de determinación espacial de los fenómenos que la ley Causa/Efecto requiere.

Pareciera que uno se encuentra en este punto en un callejón sin salida. Cualquier teoría que adscriba a los seres vivos o a un subconjunto de éstos comportamiento teleológico o que recurra a la tipo-individuación (*type-individuation* o individuación funcional) se verá enfrentada a esta paradoja. En particular, la Psicología Intencional, siendo una teoría del comportamiento humano eminentemente teleológica, presenta en sus fundamentos está aparente incompatibilidad con una concepción materialista del mundo. Muchos de sus defensores, sin embargo, serían reacios a renunciar al materialismo ontológico. Como sea, debido al objeto que persigue una teoría tal y al campo de aplicación que le corresponde, ésta no requiere de un pronunciamiento explícito respecto de este tema. En contraposición, para quienes –como yo– intentan formular una teoría de la naturalización de la intencionalidad materialista que se apoye en la Psicología Intencional el pronunciamiento respecto de esta paradoja es inexcusable.

Este fue acaso el principal escollo que se me presentó cuando comenzó a dar vuelta en mi cabeza la idea de que la intencionalidad de lo mental es una expresión de la teleología de los seres humanos. Afortunadamente para mí y también, aunque en menor medida, para los partidarios de la Psicología Intencional sin pretensiones de naturalizar la intencionalidad, di con la Teoría de la Autopoiesis de Humberto Maturana y Francisco Varela. Antes de exponer cómo esta teoría resuelve esta paradoja, haciendo A3 compatible con B, analicemos más acuciosamente la relación entre las nociones de teleología y función.

## **Función y Teleología**

Aun cuando en muchos casos puede transformarse una referencia a teleología en una referencia a función, adhiero a quienes han negado la sinonimia entre dichos conceptos

(Wright 1973). El argumento central ocupado para realizar la distinción entre éstos es que la noción de teleología (llamada también *Goal-Directedness*) como adjetivo, a diferencia de la de función, no es aplicable más que a comportamiento<sup>12</sup> de los organismos vivos. Considérese a modo de ejemplo el caso de una silla. Podemos decir sin ningún problema que la silla posee una función, toda vez que sirve para sentarse, aun cuando ésta no posee teleología, en tanto que no posee comportamiento. Podría plantearse incluso la misma cuestión respecto de ciertos órganos humanos como el corazón. Es un hecho que éste posee una función, a saber, la de bombear sangre a los tejidos corpóreos. No obstante, no resulta del todo adecuado decir que éste realiza este comportamiento (viz. bombear sangre) con el objetivo de hacer llegar la sangre a los tejidos.

En segundo lugar, podríamos plantearnos la cuestión de si todo comportamiento teleológico es en algún sentido funcional, en cuyo caso podría aceptarse que la noción de teleología, si bien no equivale a la de función, es un tipo particular de esta última. Para hacer inteligible esta relación, sería necesario establecer una equivalencia semántica entre ambos conceptos en los casos relevantes. La más plausible de éstas sería identificar los objetivos (*goals*) de un comportamiento clasificable como teleológico con el rol funcional que éste desempeña en el marco de un funcionamiento más general. Así, por ejemplo, habríamos de decir que la oración “Juan come para (con el objetivo de) alimentarse” sería traducible al lenguaje funcional como “la función de comer en Juan es alimentarse”.

Se ha dicho en este respecto que en un buen número de casos el objetivo atribuible a ciertos comportamientos teleológicos no coincide con el rol funcional que desempeñan. Considérese, por ejemplo, el caso de la copulación entre mamíferos. Claramente la función de dicho comportamiento es la reproducción de la especie, aun cuando muchas veces los mamíferos no lo hacen con este objetivo en mente –sin entrar en detalles, baste considerar el hecho de que los seres humanos usamos anticonceptivos. Otro caso en el que se ha dicho que la traducción propuesta no se sostiene es aquél donde la inmediatez de los objetivos de un cierto comportamiento no puede dar cuenta de un rol funcional bastante lejano de dicho comportamiento. Así, por ejemplo, ciertos tipos de plancton varían la distancia en que se encuentran respecto de la superficie del agua a ciertas horas del día con el objeto de mantener

---

<sup>12</sup> Incluyo dentro de la palabra “comportamiento” a los procesos biológicos. Para referirme a los intencionales ocuparé “conducta”.

la intensidad de la luz relativamente constante. Éste hecho ha sido determinado experimentalmente con luz artificial. Sin embargo, la función de este comportamiento es mantener constante la reserva de oxígeno (Wright 1973) y no la intensidad de la luz.

En estos dos casos pareciera que el comportamiento teleológico no es más que un efecto de superficie del funcionamiento de un organismo, el cual tiene relación con características más profundas enraizadas en el desarrollo evolutivo de la especie. Consecuentemente, no siempre un comportamiento teleológico puede interpretarse propiamente con la función de dicho comportamiento. La desagradable conclusión que se obtiene de éste análisis es que, de existir relación semántica alguna entre estas nociones, ésta no es la identificación del objetivo de un comportamiento teleológico con la función de dicho comportamiento. Y, sin embargo, no hay otra interpretación que parezca plausible, o, al menos, no que yo sepa.

Esta aparente imposibilidad de poner las nociones de función y teleología en una relación tan estrecha no resulta tan extraña si se considera la distinción respecto de las restricciones temporal y espacial que la ley Causa/Efecto impone sobre la determinación de los fenómenos. Mientras que la noción de teleología contraviene A3(a)(*Determinación Temporal*), la de función se opone a A3(b)(*Determinación Espacial*). Si, en cambio, ambas estuvieran vinculadas semánticamente, sería de esperarse que produjeran la misma contradicción con la ley Causa/Efecto; cuestión que, por supuesto, no ocurre.

Sean o no la teleología y la función equivalentes semánticos, es un hecho que el análisis de la última puede aportar luces respecto de la primera, en la medida que dicho análisis consiga aclarar la aparente incompatibilidad de la noción de función con A3. Con este objetivo en mente, repasaremos a continuación de manera esquemática la historia reciente de los principales esfuerzos por dar cuenta de la noción de función y las distintas dificultades que han aparecido en su camino.

Un primer tipo de teorías que se ha barajado para explicar la noción de función es la postulación de una equivalencia de ésta con alguna tipo de relación lógica. Los más habituales candidatos para esta relación han sido la de suficiencia o necesidad lógica, o bien ambas simultáneamente, llamada también equivalencia lógica. Entre estas teorías podemos mencionar la de Beckner (1959) y Nagel (1960). El primero planteó que ‘un objeto O tiene la función F en un sistema S si, y sólo si, F ocurre cuando S posee O y F no ocurre cuando S no posee O’. La primera cláusula (viz. F ocurre cuando S posee O) corresponde a decir que el

hecho que S posea O es suficiente para que F ocurra. La segunda cláusula (viz. que F no ocurre cuando S no posee O) corresponde a decir que es necesario que S posea O para que F ocurra. En resumen, el hecho que F ocurra es lógicamente equivalente, según Beckner, a la posesión de O por parte de S. Nagel, por su parte, prefirió omitir la suficiencia lógica quedándose sólo con la necesidad. Consecuentemente su postura puede resumirse como sigue: decir que la función de O es F en un sistema S equivale a decir que la presencia de O en S es necesaria para que se tenga F.

Sin embargo, algo anda mal con estas teorías que apelan a las relaciones lógicas y, en particular con las dos mencionadas. A continuación veremos dos críticas dirigidas a la concepción de suficiencia y de necesidad lógica, respectivamente, como relaciones definitorias de la noción de función. Además estará realizando crítica alguna contra la relación de equivalencia lógica como definición de función, toda vez que ésta no es más que una combinación de las anteriores.

En primer lugar, el uso de la relación de suficiencia lógica para definir la noción de función no logra distinguirla de otros posibles efectos que accidentalmente pueda causar. Considérese a modo de ejemplo el caso del sonido producido por el latir del corazón humano. Si la función del corazón quedara definida por la suficiencia de éste para producir un cierto efecto tendríamos que aceptar que la función del corazón es producir el sonido de los latidos. Más aun, la función del corazón no sería única en tanto que su actividad también es suficiente para producir otros efectos como la circulación de la sangre a través del cuerpo. Un segundo ejemplo de las dificultades que enfrentaría esta definición de función sería el caso de una persona que usara un reloj de mesa como pisapapeles en su escritorio. En la medida que dicho reloj es suficiente para sujetar papeles tendríamos que aceptar que la función del reloj es pisar papeles. No creo que alguien pudiera considerar que ésta es la función de un reloj de mesa. Consecuentemente, la suficiencia lógica no logra dar cuenta de la noción de función.

En segundo lugar, si bien la noción de necesidad lógica presenta dificultades menos evidentes que la de suficiencia para dar cuenta de la noción de función, existen variados ejemplos en los que dicha noción muestra falencias incorregibles. Considérese nuevamente el caso del corazón. Todos estaríamos de acuerdo en que la función del corazón es bombear sangre al cuerpo, no obstante, la razón de esto no puede ser la necesidad lógica, en tanto que con los desarrollos tecnológicos de las últimas décadas sabemos que el corazón no es necesario para realizar esta labor. Hoy en día existen máquinas capaces de reemplazar al corazón en el

bombeo de sangre al cuerpo y, mientras tanto, la función del corazón sigue siendo la misma. Consecuentemente, la necesidad tampoco logra explicar la noción de función.

Un segundo tipo de intento de explicar la noción de función es el de apelar a alguna forma de utilidad. Consideremos aquél llevado adelante por John Canfield (1964), quien plantea que ‘la función de un cierto objeto O en un sistema S es hacer F si, y sólo si, O hace F y el hacer F es útil para S’. A pesar de que esta explicación parece más convincente que las anteriores esta enfrenta un primer obstáculo al intentar establecer el sistema para el cual la posesión de O es útil. Por ejemplo, todos estaríamos de acuerdo en que la función del segundero de un reloj es indicar el segundo en que el observador se encuentra dado un minuto y hora determinados. La pregunta es ¿para qué sistema es esto útil? ¿Podemos decir que esto es útil para el reloj mismo? Ciertamente que no, a menos que asignemos a éste una función (o acaso una teleología) que permita realizar una afirmación tal. En otras palabras, habríamos de postular algo así como que el segundero es útil para el reloj porque le permite desarrollar su función. Pero entonces ¿de qué sistema participa el reloj que nos permite hablar de “su función”? No sé qué, pero algo huele a circularidad.

La razón de fondo que permite explicar el fracaso de la propuesta de Canfield, al menos para los intereses que perseguimos en la presente investigación, es que la noción de utilidad supone de algún modo la noción de función. O, para ser más precisos, la noción de utilidad supone, al igual que la noción de función, una relación entre hechos que no respeta A3. Decir que algo es útil invita inmediatamente a la pregunta por el fin que se persigue al hacerlo. La noción de utilidad está determinada por los objetivos de un cierto proceso. Supongamos, por ejemplo, que un niño curioso pregunta al lector si leer novelas de amor es útil. Así planteada la pregunta, no creo que alguien pueda resolver la inquietud de ese niño. Presumo –por el hecho de interesarse en el presente trabajo– que el lector diría algo así como “leer es bueno”; cuestión que, sin embargo, no responde a la pregunta. En rigor, si se quiere responder certeramente, habría que pedir a dicho niño que pusiera su pregunta en un contexto más específico. Habría que enterarse previamente del objetivo que dicho niño pretende. Una buena respuesta, según me indica el sentido común, sería: “Depende. Si lo que pretendes es ser escritor, entonces sí. Si lo que pretendes es ser futbolista, entonces probablemente no.” En general, hablar de utilidad supone hablar de una cierta pretensión o propósito. La noción de utilidad es eminentemente teleológica. Consecuentemente, explicar la noción de función en términos de la de utilidad puede que logre despejar nuestras dudas respecto de la

contradicción que supone con A3(b) (*Determinación Espacial*), pero a costa de asumir la contradicción que la noción de teleología produce con A3(a) (*Determinación Temporal*).

Como acertadamente sugiere Wright (1973), todos los anteriores análisis de la noción de función tienen algo en común: presuponen que ésta es una categoría metafísica. Tanto Nagel, como Beckner y Canfield creen estar haciendo metafísica al intentar explicar qué es para un objeto o proceso tener una cierta función. Por lo tanto, buscan incesantemente relaciones entre hechos y términos que se acomoden con la forma como habitualmente usamos la palabra. Sin embargo, la noción de función no es una categoría metafísica, sino epistémica. La adscripción de función a un cierto objeto o proceso es intrínsecamente explicativa; o, puesto de otro modo, es comprensible sólo desde la relación entre sujeto (observador) y mundo (fenómeno). Cuando alguien dice que la función del corazón es tal y tal, no está adscribiendo una propiedad intrínseca a éste, sino que lo que está haciendo es explicarse una serie de fenómenos, interconectados causalmente, bajo una estructura de análisis muy especial, distinto del que habitualmente se ocupa en disciplinas como la Física. Se está dando preponderancia a una de las propiedades causales que posee el corazón (viz. la de bombear sangre) que son relevantes para el observador por las razones que sean, por sobre otras propiedades (e.g. la de latir).

Según creo, la teleología y, en general, toda atribución al mundo de un orden de determinación de los hechos que no se rige por A3, debe tener carácter explicativo. Esta es la única forma de evitar la contradicción con A3, pues, en este caso el materialismo y la teleología, al ser un postulado metafísico el primero y epistémico el segundo, no compiten por un mismo espacio. El materialismo constituye, pues, una propiedad del mundo, la teleología – como también la noción de función- es parte del modo como nosotros, los hombres, nos relacionamos con (interpretamos) ese mundo.

Cada cosa en su lugar y podemos estar tranquilos. Sin embargo, esta simple declaración no agota el problema, sino que abre otro tan grande como el anterior que, hasta donde yo sé, no ha recibido suficiente atención por parte de la Filosofía Analítica. La enorme pregunta que surge es ¿cómo es metafísicamente posible que la explicación por medio de las nociones de teleología y función, que invierten el orden de determinación de los hechos, funcionen –valga la redundancia- en un mundo regido por la ley Causa/Efecto (A3)? La teoría de la Autopoiesis nos permitirá esbozar una respuesta.

## **Autopoiesis**

Talvez no estaba entre los objetivos principales de Maturana y Varela (1964) dar cuenta de la teleología de los organismos vivos; pero lo hicieron. Y lo hicieron, por supuesto, asignando un rol fundamental en la explicación al observador. La teleología no puede ser una categoría metafísica, a menos que queramos renegar del materialismo. La gracia debe ser, pues, explicar la vida en términos mecanicistas y, más aun, hacerlo de modo que atribuir teleología a los organismos vivos no sólo sirva como explicación de su comportamiento sino que sea además atractiva como explicación del fenómeno por el observador. Aquí justamente radica la gran potencia de la Teoría de la Autopoesis.

Para exponer adecuadamente la teoría, es necesario partir por la noción de máquina homeostática. Una máquina, en primer lugar, es cualquier sistema material complejo que posee una dinámica interna y tiene unidad espacial y temporal. Una máquina homeostática, a su vez, es una máquina cuya actividad mantiene una variable constante en el tiempo. Así, por ejemplo, una estufa con termostato es una máquina homeostática toda vez que su actividad mantiene constante la variable “temperatura ambiental”. No estamos diciendo con esto que la función de la estufa sea mantener la temperatura ambiental inalterable o que funcione para conseguir este objetivo, sino que simplemente constatamos el hecho. Más aun, si es que habitualmente asignamos esta función a dicha máquina es porque de hecho mantiene la temperatura ambiental constante, característica que fue explícitamente considerada en su diseño. Consecuentemente, siempre podemos describir la actividad de la máquina sólo apelando a relaciones causales que respetan la determinación de los fenómenos exigida por la ley de Causa/Efecto, aun cuando como observadores generalmente adoptamos la explicación funcional.

Una máquina autopoietica será, pues, definida, según Maturana y Varela, como un tipo de máquina homeostática; específicamente, como una máquina cuya actividad mantiene constante la variable “organización interna”. Del mismo modo como la actividad de una estufa con termostato mantiene constante la temperatura ambiental, la actividad de una máquina autopoietica no altera la organización interna de la máquina misma. Obviamente, al igual que en el caso de la primera, no estamos diciendo que la función de la máquina autopoietica sea mantener constante su estructura, sino que nuevamente planteamos que de hecho es así.

Es importante notar que una máquina tal es altamente dinámica y muy sensible a perturbaciones de su entorno. Consecuentemente, ésta está constantemente intercambiando materia y energía con su entorno. De este modo, parte de la actividad misma de la máquina autopoietica es estar persistentemente produciendo (regenerando) sus propios componentes manteniendo así su organización interna constante. Ésta es otra forma de definir lo que es una máquina autopoietica, a saber, como una que produce sus propias componentes y, que, por lo tanto, se produce a sí misma –de ahí viene su nombre. Nótese nuevamente la meticulosa neutralidad de esta definición respecto de la noción de función. No hemos dicho que una máquina autopoietica sea una cuyo producto o salida (output) sea ella misma, sino que hemos dicho que de hecho se produce a ella misma sin perjuicio de que pueda producir otras cosas (sonidos, calor, movimiento, otras máquinas autopoieticas, etc.). Aun así, son muy pocas las cosas que una máquina autopoietica produce que no participan del proceso mismo de autopoiesis. Sonidos, movimiento o toda materia que de ella salen pueden entenderse en la mayor parte de los casos como parte de los procesos mismo de producción interna. Esto es lo que hace que puedan ser entendidas funcionalmente como sistemas cerrados o, al menos, como no teniendo salidas (outputs) al exterior.

A modo de explicación, podemos considerar una máquina autopoietica como una que, frente a frecuentes perturbaciones de su entorno, realiza cambios compensatorios que permiten que ésta mantenga su organización interna inalterable. Por supuesto, no obstante suelen ser muy eficientes en la ejecución de ésta actividad, las máquinas autopoieticas no son infalibles, de modo que existirán ciertas perturbaciones del entorno frente a las cuales perderá su organización, dejando así de realizar los procesos de producción de sí misma y dejando, en definitiva, de ser autopoieticas. En palabras de Maturana y Varela, existe un dominio de perturbaciones frente a las cuales una máquina tal podrá mantener su actividad, fuera del cual se romperá este círculo virtuoso de auto-producción.

Como sea, la misma característica autopoietica de dichas máquinas las provee de una identidad. La producción de sus componentes (o auto-producción) es una tarea fundamental en su funcionamiento, y no gracias a la estipulación del observador, sino que al hecho que estas máquinas destinarán la mayor parte de la energía que capten de su entorno en realizarla. Podría decirse según Maturana y Varela que una máquina autopoietica es una que transforma energía en sí misma. Así, como casi toda su producción material es parte de ella o del proceso mismo de autopoiesis, esto define muy marcadamente un dominio de fenómenos espacial y

temporal. Esto, a su vez, permite que se establezca como sistema de procesos auto-referente y que tenga así una identidad fenoménica muy marcada.

Pues bien, la idea de definir máquinas autopoieticas es decir al mismo tiempo que éstas coinciden exactamente con lo que entendemos como vida. Esto gracias a que una definición tal de vida explica con bastante precisión todos los fenómenos que se observan en los organismos a los que habitualmente se les atribuye dicha condición (Maturana & Varela 1964). Pero eso no es todo. Lo más importante en relación con la presente investigación es que dicha noción de los organismos vivos es perfectamente compatible con la ley Causa/Efecto. Esto se hace patente en el hecho que nunca recurrimos a las nociones de teleología o función para definir autopoiesis, sino sólo a una descripción fenoménica cuyo orden de determinación de los hechos es el requerido por dicha ley. Más aun, no es necesario recurrir a nociones tales para describir el funcionamiento de cualquier máquina homeostática (e.g. una estufa con termostato), siempre y cuando se suponga respecto de ésta su calidad ontológica de unidad. Basta considerar dichos procesos como cadenas causales restringidos al espacio físico que dicha unidad supone y que terminan en ciertos resultados como la producción de la máquina misma, sin perjuicio de que puedan haber también otros (como la producción de calor en el caso de los mamíferos). No hay ningún resultado que tenga ninguna prioridad sobre otro, salvo por el hecho que generalmente una máquina autopoietica invierte la mayor parte de la energía que capta del entorno en su auto-producción.

Alguien podría replicar que la apelación al concepto de máquina supone una forma de individuación funcional, frente a lo cual habría que responder que –como observa Lycan– ninguna forma de individuación (en consecuencia, ninguna ontología) puede escapar a este hecho, salvo que nos refiriéramos directamente a la materia, sin reconocer ninguna propiedad en ella –exigencia que ni la Física podría tolerar (Lycan 1987). Esta explicación de la vida puede jactarse, sin embargo, de ser una tipo-individuación (*type-individuation*) en su grado mínimo, toda vez que no recurre a ninguna forma de determinación de los fenómenos que no sea la permitida por la ley Causa/Efecto, salvo, acaso, en el reconocimiento de la unidad ontológica de la máquina, cuestión que, nuevamente, depende de la ontología que reconocemos.

Pero la autopoiesis no sólo da una explicación del fenómeno de la vida en términos mecanicistas, sino que además permite comprender porqué la atribución de teleología por parte del observador funciona como explicación del comportamiento de este tipo de

organismos y –igualmente importante- porque frecuentemente resulta más eficaz que la explicación causal. Respecto de lo primero el punto clave es que si casi toda la energía que una máquina autopoietica capta del entorno se destina a la producción de ella misma, entonces resulta sumamente natural decir que casi todo lo que una máquina autopoietica hace lo hace *para* mantener su condición viva (o autopoietica), toda vez que las consecuencias en términos de predecir el comportamiento de una máquina tal son las mismas en uno y otro caso. Así, por ejemplo, uno puede inferir exactamente lo mismo respecto del comportamiento de un cierto animal cuando es atacado (viz. que seguramente se defenderá) atribuyéndole autopoiesis o diciendo que actuará *para* resguardar su integridad. En otras palabras, se extraen las mismas inferencias al decir que un cierto organismo es autopoietico que al decir que su comportamiento tiene la conservación de su condición autopoietica como objetivo (*goal*). Más aun, como observa Dretske a instancias de Gallistel, la teleología (*Goal-Directedness*) como explicación se expande a todo comportamiento que se supedita a un objetivo principal y, por lo tanto, casi todas las conductas de una máquina autopoietica se podrán entender teleológicamente. Más adelante diremos más sobre esto último. De momento podemos identificar la auto-producción de este tipo de máquinas como el fin último o *teleos* aristotélico.

En segundo lugar, está el hecho que el observador situado frente a una máquina autopoietica generalmente preferirá la explicación teleológica. Esto se explica en base a dos hechos. En primer lugar, la noción de autopoiesis es un término teórico que resulta demasiado complejo para ser usado cotidianamente, mientras que la noción de teleología es de manejo habitual entre los seres humanos –baste considerar lo que se ha llamado la Psicología Popular. Pero además, existe una razón muy pedestre aunque no menos convincente para explicar este hecho, y es que la actividad interna de las máquinas autopoieticas (u organismos vivos) no está usualmente a la vista del observador. Por el contrario, normalmente el observador común (viz. no científico) no cuenta con ninguna pista que le permita realizar hipótesis respecto de lo que ocurre al interior de una de estas máquinas, a la vez que la explicación teleológica no requiere de ninguna hipótesis respecto de dicho funcionamiento, sino tan sólo de una relación entre entradas (inputs) y salidas (outputs). Podríamos incluso pensar que el recubrimiento (piel, caparazón, membrana, etc.) que normalmente exhiben los organismos vivos es una componente importante en su eficacia para mantener su organización constante. No tenerlo los haría más vulnerables a las perturbaciones de su entorno.

Finalmente quisiera hacerme cargo de una posible crítica que he escuchado respecto de la noción de autopoiesis como definición de lo vivo. Alguien podría decir:

“Al esforzarse por excluir cualquier noción funcional en dicha definición – como sería decir que la función de una máquina autopoietica es la de auto- producirse o que la salida (output) de ella es ella misma- se es demasiado *liberal*. O, en otras palabras, que muchas máquinas de las cuales sabemos que no constituyen vida (e.g. un automóvil) caerían bajo la definición, en tanto que su actividad mantiene constante su organización interna.”

Frente a lo cual cabría dar dos respuestas. En primer lugar, que las máquinas autopoieticas no sólo se definen por mantener su organización constante, sino que además por ser altamente dinámicas y, consecuentemente, por captar constantemente energía de su entorno transformándola en sus componentes. Un automóvil ciertamente pasa gran parte del tiempo en una situación estática y, adicionalmente, no produce sus propios componentes.

En segundo lugar, tendríamos que decir que la actividad de un automóvil mantiene su organización constante tanto como mantiene la temperatura ambiental constante. No es sólo que una cierta máquina haga algo sino que lo realmente importante es cuan efectiva es para hacerlo frente a las condiciones relevantes. En otras palabras decir qué hace y qué no hace una máquina es un asunto de análisis contrafáctico. Puede ser que bajo ciertas condiciones ambientales un automóvil mantenga la temperatura de un cierto espacio físico constante, pero es también claro que rara vez se dan esas condiciones. Del mismo modo, puede que un automóvil mantenga su organización constante bajo ciertas condiciones (e.g. mientras se está moviendo), pero por cierto que no lo hace frente a muchas otras en que las máquinas autopoieticas sí (e.g. frente a la presencia de un aprendiz de mecánico ávido de experiencia).

## **Psicopoiesis**

Según ha planteado la Psicología Intencional, el hombre actúa en pos de la consecución de objetivos. Esto es lo que se quiere decir cuando se postula que la adscripción de intenciones a los hombres es la clave para entender su conducta. Sin embargo, -más allá de una discusión semántica detallada- para tener intenciones hay que tener antes actitudes proposicionales (deseos y creencias). Así, por ejemplo, cuando se explica la conducta criminal de un ladrón de bancos en términos de sus intenciones de apoderarse de un suculento botín, se requiere a la vez la adscripción a dicho delincuente de, al menos, el deseo de tener dinero y la creencia de que asaltando dicho banco lo podrá conseguir. En otras palabras, toda intención puede

explicarse en términos de deseos y creencias por parte del sujeto. En este sentido, hablar de conducta intencional es al mismo tiempo hablar de actitudes proposicionales.

De este modo, no resulta raro decir que las actitudes proposicionales son el motor de la conducta teleológica. Esto es, que previo a la realización de una cierta acción (dejando a un lado las respuestas reflejas), el ejecutor de ésta debe tener ciertos deseos y creencias. Sin estos no hay conducta, o, al menos, no hay conducta intencional. Esto es precisamente lo que permite individuar a los deseos y las creencias: el rol causal (funcional) que éstas tienen en la actividad mental y, en último término, en la conducta<sup>13</sup>. Más aun, en la medida que la teleología de los organismos vivos constituye una característica más profunda que el sustrato mental, las actitudes proposicionales son precisamente la expresión de aquélla en los organismos dotados de mente. Las actitudes proposicionales, propias de los organismos cerebrados, constituyen, pues, la posibilidad de elevar la autopoiesis de estos organismos al nivel mental.

Los organismos cerebrados (viz. con sistema nervioso central) son también organismos vivos y, por lo tanto, máquinas autopoieticas. Postular que el cerebro es un órgano integrado a la autopoiesis de dichas máquinas permitirá no sólo entender el rol que éste juega en el funcionamiento biológico de ellas, sino que también nos permitirá entender qué es la actividad mental en términos de su rol causal en la conducta. En pocas palabras, estamos tratando de establecer una cuestión que puede sonar evidente pero que nunca está demás explicitar: la mente hace el proceso de adaptación al entorno más eficiente. O, evitando la noción de adaptación spenceriana, la mente hace el proceso de autopoiesis de los organismos vivos contrafácticamente más eficiente (viz. existe un dominio más amplio de perturbaciones del entorno frente a las cuales la máquina podrá conservar su autopoiesis). Y las actitudes proposicionales, por su parte, son la expresión de la autopoiesis en las conductas intencionales o mediadas por la mente (excluyendo ciertas respuestas puramente biológicas).

Pero el objetivo de la presente tesis es mostrar cómo, a partir de la explicación de la vida por vías de la autopoiesis, se comprende toda su teleología y, especialmente, la teleología asociada a la conducta intencional. Para esto habremos de considerar un hecho muy significativo, a saber, que la teleología que habitualmente atribuimos a casi todas las

---

<sup>13</sup> Me refiero a los deseos y creencias a secas y no a los *deseos-que-P* o *creencias-que-P*. Ver Fodor, J. 1987, Cap. 3

conductas (intencionales o no) que exhibe un organismo vivo son una forma que tiene el observador de interpretar el proceso de auto-producción que define al organismo como unidad en términos de una estructura jerárquica de explicaciones. En otras palabras, existe una correspondencia biunívoca entre lo que entendemos como conducta teleológica y subproceso autopoiético; de modo que toda conducta teleológica se enmarca dentro de un red jerárquica de fines particulares (*goals*), cuyos ítems corresponden a un subproceso autopoiético y cuyo primer eslabón (o eslabón superior) –que podría identificarse con el *teleos aristotélico*– correspondería, a su vez, con el proceso autopoiético completo.

En el caso de los organismos dotados de mente esto es particularmente notorio. Consideremos el famoso ejemplo usado por Dretske (1991) para recalcar este punto. Clyde está sentado en su sillón y de pronto siente deseos de beber una cerveza. Se levanta del sillón, ayudándose con sus manos y comienza a caminar hacia la cocina porque cree que allí encontrará una lata de cerveza. En su camino se encuentra con su perro que está echado en el suelo, y lo elude para finalmente entrar a la cocina. Una vez adentro utiliza su mano derecha para abrir el refrigerador y levanta la mano izquierda, abre los dedos y finalmente alcanza la lata deseada.

Todo este proceso que nace en el deseo de Clyde de beber una cerveza, es altamente complejo y sería inoperante intentar detallarlo exhaustivamente. Sin embargo, podemos suponer sin pérdida de generalidad que la descripción entregada considera todas las conductas involucradas en él. El hecho interesante es que la descripción teleológica de este comportamiento supone una red jerárquica de objetivos<sup>14</sup> que lo explica, la cual tiene en sus extremos inferiores las respuestas motoras, cotidianamente entendidas como conductas.

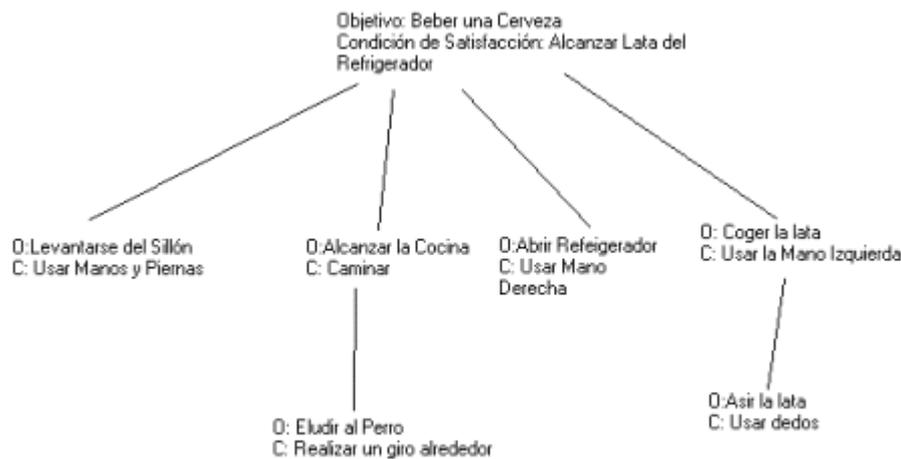
Consideremos el esquema de esta red ilustrado en la figura 1. Cada objetivo en esta estructura da sentido a los objetivos que al mismo tiempo subsume. En esta estructura jerárquica radica la naturaleza misma de la explicación teleológica de la conducta. A su vez, cada objetivo es expresable en lenguaje intencional por medio de la adscripción de un deseo a Clyde, mientras que cada condición de satisfacción puede entenderse por medio de la adscripción de una creencia respecto del modo de alcanzar el objetivo. No debe confundirse, sin embargo, las condiciones de satisfacción que estamos usando aquí, las cuales son *a posteriori*, con las condiciones de satisfacción que un deseo posee de manera intrínseca. Esto es, la condición de

---

<sup>14</sup> Dretske le llamaría “estructura de control jerárquicamente organizada”, siguiendo a Gallistel 1980.

satisfacción intrínseca del deseo de beber cerveza es, precisamente, beber cerveza. Sin embargo, dadas las condiciones en que Clyde se encuentra, esto se obtiene tomando aquella que está en el refrigerador, o, al menos, eso cree él. Esta última es la condición de satisfacción a la que nos referimos aquí. Todos los objetivos adquieren, pues, sentido en virtud del objetivo que los anteceden en esta estructura jerárquica. Todos excepto el primero (viz. beber una cerveza). La razón es que hay otros objetivos implícitos en la descripción de la conducta de Clyde. Falta en la descripción anterior una rama completa de la red jerárquica ubicada en lo más alto, que da sentido a todo lo que sigue. Así, por ejemplo, podríamos decir que Clyde desea tomar cerveza *para* hidratar su cuerpo, lo cual supone un objetivo superior de carácter biológico (viz. no intencional), como muchas veces –aun cuando no necesariamente- habrá de ocurrir. Sin embargo, lo que es seguro, y esto es lo relevante para nuestro análisis, es que en la cima de toda esta estructura, la explicación no puede ser intencional ni siquiera biológica propiamente tal, sino que debe ser autopoietica. En otras palabras, la autopoiesis del organismo constituye el objetivo final o *teleos aristotélico* y da sentido a toda la estructura que lo sigue, incluida aquella descrita intencionalmente.

Figura 1



Cada nivel de la descripción biológica en la estructura de comportamiento de Clyde corresponde a un subproceso autopoietico que no está mediado por la mente, mientras que cada nivel de la descripción intencional corresponde a un subproceso autopoietico que sí lo está. La comprensión del primer tipo de correspondencia (viz. descripción biológica/subproceso autopoietico) no reviste mayor dificultad, toda vez que la noción central

de la descripción biológica es la de función, la que remite directamente al funcionamiento del todo como sistema, que es precisamente la actividad autopoietica<sup>15</sup>. En la mediación de la mente está, pues, la gran dificultad de entender la autopoiesis de los organismos que están dotados de ella. Intentaremos, pues, a continuación, esbozar una explicación de qué es la mente desde una perspectiva funcionalista, como se ha puesto de modo desde las críticas de Putnam, Fodor y Block contra la Teoría de la Identidad. La novedad radicará en el hecho que, en este caso, la función de la mente se determinará en términos de una concepción del organismo que la posee como máquina autopoietica, y que tiene como salida (output) el conjunto de procesos que la componen<sup>16</sup>.

Para comprender cómo participa la mente de la autopoiesis de los organismos que la poseen consideremos tres hechos fundamentales respecto de ella:

- a) La mente coordina un gran número de procesos del organismo.
- b) La mente, de algún modo, considera constantemente la situación externa para realizar su labor.
- c) La mente tiene una influencia casi nula en la mayor parte de los procesos biológicos.

La cláusula (a) da cuenta del rol de la mente como productora de subprocesos autopoieticos específicos organizados en una estructura jerárquica<sup>17</sup> y recursiva similar a la expuesta en el ejemplo de Clyde, de modo que cada rama de la estructura se realiza con independencia de las otras, aun cuando todas juntas resultan en el ajuste o reparación de un aspecto específico del funcionamiento y, finalmente, en conservación de la organización interna de la máquina. Así, por ejemplo, en el caso de Clyde la segunda rama de la descripción teleológica de su conducta (viz. “alcanzar la cocina/eludir al perro”) se realiza con independencia de la cuarta (viz. “coger la lata/asir la lata”), sin perjuicio de que ambas sean parte de un proceso mayor que

---

<sup>15</sup> Más sobre la descripción biológica en términos de la autopoiesis en Maturana & Varela 1964.

<sup>16</sup> Adoptaremos un funcionalismo teleológico en el sentido que lo dice Sober 1985, toda vez que disponemos de una forma de entender esta noción en términos de la de autopoiesis.

<sup>17</sup> En palabras de Dretske (1991) “estructura de control”.

resulta en la satisfacción del deseo de beber cerveza de Clyde y, posiblemente, en la hidratación de sus tejidos corporales.

La cláusula (b) da luces respecto del funcionamiento interno de este aparato coordinador de actividades y, al mismo tiempo, nos puede ayudar a distinguir los roles funcionales de las creencias y deseos en la autopoiesis de los organismos dotados de mente. Así, podemos pensar que el nacimiento de cada subproceso a partir de otro que lo antecede en la estructura, requiere, a su vez, de una evaluación de las condiciones externas en las que desenvuelve el organismo. Por ejemplo, cuando Clyde camina hacia la cocina, recibe un estímulo visual que le indica que su perro está en el camino entre él y su destino, de donde nace un nuevo subproceso que lo llevará a eludirlo. Usualmente, aunque no necesariamente (e.g. la creencia de que hay una lata en el refrigerador), la evaluación de las condiciones externas será hecha en base a estímulos externos inmediatos. Por otro lado, la evaluación de las condiciones exteriores es lo que llamamos creencia, y su rol funcional no es la producción misma de subprocesos sino que, más bien, la determinación de qué nuevo subproceso se iniciará. Los deseos, por su parte, serán los encargados de poner en marcha los procesos determinados por las creencias. Metafóricamente, podemos entender los deseos como la fuerza motora de la conducta, mientras que las creencias como el manubrio que guía el desenvolvimiento de ésta.

Finalmente, la cláusula (c) nos indica qué tipo de procesos son aquéllos mediados por la mente. A las características ya mencionadas de estos procesos como su coordinación y relación con el entorno, se suma otra tan relevante como éstas. Los procesos autopoieticos mediados por la mente no producen resultados biológicos inmediatos, es decir, no producen resultados respecto de la actividad interna del organismo, sino que, más bien, su campo de acción parece estar restringido a la relación de la unidad autopoietica con su entorno circundante. Esto en ningún caso quiere decir que estos procesos no participen de la autopoiesis del organismo; lo hacen, pero su aporte consiste en la producción de relaciones de éste con su entorno que juegan un rol en el adecuado funcionamiento de los procesos autopoieticos internos y, con ello, en la conservación de la organización interna. Esto permite llamarlo autopoieticos, aun cuando no produzcan componente alguna del organismo. Más aun, hay un hecho no menos significativo respecto del rol de la mente en la autopoiesis como responsable de la relación del organismo con su entorno y es que su actividad se traduce en cambios en la disposición de aquél respecto de objetos *específicos* que lo rodean. En el ejemplo de Clyde, los deseos de beber cerveza, eludir al perro, etc. y las creencias de que hay una lata de cerveza en el refrigerador, de que su perro está echado en el pasillo, etc. producen

cambios en la disposición de Clyde respecto específicamente de la lata de cerveza, el refrigerador, el perro, etc. A este tipo de subprocesos autopoiéticos con esta característica que llamaré *especificidad* y que tienen impacto, aunque indirecto, en el funcionamiento interno, los denominaremos en adelante procesos psicopoiéticos. Su nombre se debe a que sólo la mente es capaz de producirlos.

Pero eso no es todo. Acaso la principal propuesta de este capítulo no es la mera existencia de los procesos psicopoiéticos sino el que éstos constituyen precisamente lo que cotidianamente llamamos conducta (técnicamente, conducta intencional)<sup>18</sup>. No es fácil medir que tan buena es esta definición, toda vez que la noción misma de conducta es bastante laxa y difícilmente podemos encontrar una definición precisa e independiente que permita establecer una comparación<sup>19</sup>. En una aproximación de primer orden, podemos decir que la noción de conducta como psicopoesis al menos respeta la idea intuitiva de que aquello que entendemos como tal siempre tiene como causa inmediata estados mentales. Pero eso no es suficiente. Acaso la única forma de establecer la calidad de la identificación de psicopoesis con conducta será por medio del análisis de la capacidad de la primera para dar cuenta de los fenómenos que clasificamos de manera pre-teórica como conducta, especialmente en los casos límite. Y, según creo, la noción de psicopoesis satisfará los requerimientos. Así, uno de los casos que mayormente incomodó a los conductistas radicales fue aquél en que un individuo procuraba mantenerse completamente quieto frente, por ejemplo, a la presencia de un brutal asesino que lo acechaba en la oscuridad. Al no haber respuesta motora dicha conducta por parte de la víctima era difícilmente clasificable como tal por los conductistas, viéndose obligados a recurrir a nociones como “disposición” al movimiento u otras, que no encajaban armónicamente con su teoría. Bajo nuestra noción de conducta como psicopoesis, sin embargo, ese tipo de fenómenos es naturalmente clasificable como conducta. El individuo permanece quieto pues este es el resultado de un proceso mental que termina con el establecimiento de una relación entre la máquina autopoiética (víctima) y el asesino, que tiene como consecuencia la conservación, al menos momentánea, de la autopoesis de aquélla.

---

<sup>18</sup> Prefiero hablar de conducta intencional antes que de acción, puesto que esta última noción ha sido habitualmente asociada con la noción de voluntad, la cual merecería un tratamiento más detallado.

<sup>19</sup> Más sobre esto en Dretske, F. 1991.

Todas las anteriores virtudes que se pueden asignar a la noción psicopoiesis provienen, sin embargo, de una sola gran virtud, a saber, la capacidad de fundar por medio de la noción de autopoiesis una teoría funcionalista teleológica de la mente. Como veo las cosas, la diferencia fundamental entre el Funcionalismo Teleológico y el Funcionalismo Máquina de Turing (Sober 1985) es que el primero es más restrictivo que el segundo, puesto que recurre a la individuación global y no local, como lo hace el segundo. En otras palabras, la noción de función implicada en el Funcionalismo Máquina de Turing supone que, por ejemplo, un estado mental ha de individuarse en términos de lo que lo produce (input) y de lo que éste causa (output) sin más, mientras que en la primera esto no basta sino que se requiere además que dichos inputs y outputs sean parte de un proceso global. Esta es, según creo, la noción de función a la que apela Sober (1985) cuando plantea que la función del corazón como bombear sangre es el modelo a seguir para individuar los estados mentales. Y, como muestra el mismo Sober, un Funcionalismo Teleológico consigue evitar muchas de las dificultades en que incurre un Funcionalismo Máquina de Turing. Más adelante volveremos sobre algunas de estas dificultades. De momento convengamos que talvez la gran fuerza que ha sostenido el Funcionalismo Máquina de Turing es la enorme dificultad con que se han topado los filósofos de la mente (darwinistas incluidos) al intentar dar cuenta del proceso global que permitiría establecer un funcionalismo teleológico naturalizado.

El Darwinismo, gran esperanza de quienes defienden un funcionalismo teleológico, ha fracasado por una razón bastante sencilla, y es que la Teoría de la Evolución supone de alguna manera la noción de vida. La noción misma de especie supone un cierto conjunto de organismos *vivos*, sin ella la teoría de la evolución queda sin nada que explicar. Se requiere, en consecuencia, una definición independiente de qué es la vida. Y la vida es, por otro lado, el proceso global que necesitamos para establecer un funcionalismo teleológico. No concibo organismos teniendo estados mentales que no estén al mismo tiempo vivos. Pues bien, la teoría de la autopoiesis constituye una definición independiente de vida y, más aun, una que sólo supone un mundo material y mecánico. A través de ella podemos establecer qué es la mente funcionalmente, al estilo Sober, y al mismo tiempo podemos decir cuál es el proceso global por medio del cual individuaremos los estados mentales: la psicopoiesis. He aquí el propósito principal de la presente investigación.

En resumen, introducir la teoría de la autopoiesis al estudio de la mente puede ser de tremenda utilidad y permite, a su vez, estudiarla de un modo funcionalista y teleológico sin miedo de caer en una abdicación del materialismo ontológico. La noción misma de mente puede ser

entendida así con mayor claridad como un órgano coordinador de procesos y subprocesos específicos (viz. focalizados en objetos específicos) cuyo resultado directo no es la adecuación de las condiciones internas de funcionamiento autopoietico sino la adecuación de la disposición del organismo en relación con su entorno circundante. Llamaremos a este tipo de procesos psicopoesis y los identificaremos con nuestra noción cotidiana de conducta intencional. Por otro lado, la forma como actúa este órgano coordinador procesos que llamamos mente es por medio de dos herramientas fundamentales como son los deseos y las creencias. Los primeros se individualizan funcionalmente como aquellos capaces de poner en marcha los procesos psicopoiéticos mismos, mientras que las segundas actúan determinando qué procesos se llevarán a cabo. Todo esto, sin embargo, se enmarcará en una estructura jerárquica de procesos y subprocesos más amplia que tendrá, comúnmente, procesos biológicos (internos) antecediendo y sucediendo a los anteriores en la jerarquía y que en una descripción teleológica darán sentido a éstos en términos de un objetivo final como es la conservación de la organización interna o autopoiesis.

## **Otras razones filosóficas para creer en la Autopoiesis**

Ya hemos visto que entender la vida por medio de la Teoría de la Autopoiesis de Maturana y Varela trae consigo varios beneficios, como son la explicación de la teleología, una visión coherente de la actividad mental, y una definición plausible de la conducta. Si aun así el lector no está suficientemente convencido de la potencia explicativa de esta teoría en el ámbito de la ciencia cognitiva, en lo que resta, creo, terminará estándolo. En efecto, a continuación nos ocuparemos de aplicar dicha concepción de la vida y la actividad mental a la resolución de dos problemas clásicos en la filosofía de la mente de los últimos 30 años. Estos son: (a) el problema del Homunculismo y (b) El problema del Chauvinismo versus Liberalismo<sup>20</sup>. Ambos son representantes de un conjunto mayor de problemas, todos relacionados con las dificultades para explicar la naturaleza de la teleología.

a) El Problema del Homunculismo: Dennett (1975) describe la metodología que sigue un investigador de Inteligencia Artificial del siguiente modo:

El Investigador de IA comienza con un problema caracterizado intencionalmente (e.g. cómo consigo que un computador *entienda* preguntas en inglés?), lo divide en sub-problemas también caracterizados

---

<sup>20</sup> También conocido como el problema de la Población de China.

intencionalmente (cómo consigo que un computador reconozca preguntas, distinga sujetos de predicados, ignore las cuestiones irrelevantes?) y luego vuelve a dividir estos problemas hasta que finalmente alcanza problemas o descripción de tareas que son obviamente mecánicas.

Aplicar la misma estrategia de investigación para entender la psicología humana parece inevitable. A partir de una descripción intencional de la conducta humana, procedería una descomposición sucesiva de ésta en sub-tareas hasta el punto en que nos encontraríamos en un nivel descriptivo puramente mecánico, probablemente neurofisiológico. Consecuentemente, las entidades que participarían de las tareas y procesos de cada nivel serían también descompuestas en sub-entidades sucesivamente, hasta llegar al nivel de descripción neuronal o, en principio al menos, al molecular o atómico. Las entidades en que se descomponen aquellas del nivel superior en la descripción se han llamado homúnculos (*homunculi*). Esta estrategia de descomposición se conoce, pues, como Homunculismo.

El Homunculismo ha sido adoptado mayormente dentro de las corrientes funcionalistas de la ciencia cognitiva. Esto pareciera indicar que las entidades mentales individuadas funcionalmente (e.g. actitudes proposicionales) debieran poderse descomponer en otras sub-entidades hasta el nivel neuronal. Sin embargo, como observa Lycan, en la descomposición de entidades individuadas funcionalmente en otras sub-entidades nunca se podrán eliminar las referencias a teleología o función (Lycan 1987). Cada vez que hablemos de una entidad E en un estado S, la individuación de éstos se hará por medio de sus poderes causales relevantes para la producción de un cierto fenómeno o proceso asociado a entidades que los mismos componen. Aun cuando aceptemos que toda instanciación de un estado mental es idéntica a un estado cerebral, y aun cuando podamos descomponer todo estado cerebral en una estructura neuronal; aun así, hay allí una referencia implícita a función en la alusión a neuronas. La individuación de éstas entidades como neuronas es sólo contingente, así como también lo es la individuación del aparato donde se instancia el estado mental como cerebro. En realidad, en lugar de neuronas podría estar cualquier entidad material que tuviera las propiedades adecuadas que permitieran las interacciones causales correctas; o, en otras palabras, que cumpliera la función esperada. Más aun, si llegásemos al nivel de descomposición atómico, habríamos también de aceptar la posibilidad lógica de que en lugar de átomos hubiera otras entidades que tuvieran las características que en los átomos son relevantes para producir los procesos descritos en el nivel superior de modo funcional. Resulta, sin embargo, que sabemos que, contingentemente, las únicas entidades materiales

que hay en el mundo son átomos, en vista de lo cual ni siquiera contemplamos esa posibilidad.

Por lo tanto, el problema del homunculismo puede plantearse así: ¿si la noción de función no es compatible con A3 (arriba), entonces adoptar la descripción funcional supone *ipso facto* renunciar al materialismo (específicamente a la ley Causa/Efecto) sin importar cuan lejos lleguemos en la descomposición homuncular? Mi respuesta es no. No necesariamente el funcionalismo es incompatible con una concepción materialista del mundo. Lo que ocurre es que, como se ha tratado habitualmente el Funcionalismo, no hay claridad respecto del proceso completo a describir ni de cómo éste se inserta en un mundo regido por la ley Causa/Efecto. Se ha dicho que una actitud proposicional es un tipo estado mental que juega un rol causal en la conducta. Pero nadie ha dicho con precisión qué tipo de fenómeno físico constituye una conducta. Puestas así las cosas no parece haber nexo con el mundo material. La noción de psicopoesis, sin embargo, da respuesta a dicha pregunta entregando una definición plausible de qué es una conducta como un tipo de subproceso autopoietico. Más aun, lo hace de un modo perfectamente compatible con una concepción materialista del mundo aun cuando suponga una forma de tipo-individuación de la vida. Según hemos dicho, un organismo vivo no es un cierto conjunto de partículas A puestas de tal y tal forma, sino es un cierto tipo de proceso físico (autopoiesis) implementable en cualquier sistema material que tenga las propiedades adecuadas. Pero no sería correcto decir que esta definición de autopoiesis supone una abdicación del materialismo, por la misma razón que hablar de derrumbes tampoco lo supone, sin importar que un derrumbe pueda implementarse en distintos objetos materiales que tengan las propiedades adecuadas (piedras, latas de conservas en un supermercado). Esto se debe a que toda instanciación de derrumbe como de autopoiesis está regida por la ley Causa/Efecto. Del mismo modo, toda instanciación de conducta, entendida psicopoiéticamente, está regida por esta ley.

En resumen, hablar de conducta o de actividad mental es al mismo tiempo hablar de vida. Ningún subproceso biológico, intencional o autopoietico en general, tiene nada especial como proceso físico excepto porque participa del proceso completo de la autopoiesis. Estudiar la mente es estudiar al hombre y estudiar al hombre es estudiar la vida, por lo tanto, éste es el ámbito relevante de análisis: la autopoiesis como fenómeno físico peculiar. La vida define un dominio de análisis especial dentro de un universo regido por la ley Causa/Efecto fuera del cual todo subproceso autopoietico no pasa de ser un mero fenómeno físico.

b) El problema del Chauvinismo versus Liberalismo: Durante la década del 70 Davis, Dennett y Block, entre otros, pusieron en cuestión el Funcionalismo al presentar las dificultades que implicaba ser demasiado liberal al definir qué es un estado mental. Si, por ejemplo, individuamos una neurona sólo por su rol causal en una estructura, entonces, literalmente, casi cualquier cosa puede ser una neurona. Consecuentemente, si consideramos un gran grupo de personas con un martillo en su mano contratadas para situarse en una disposición isomórfica a las neuronas de un cierto estado mental (e.g. ‘desear un helado’), de modo que cada vez que una neurona manda una señal a otra, la persona correspondiente golpea a la otra en la cabeza con el martillo; entonces tendríamos que aceptar que el conjunto de personas, sea lo que esto sea, desea un helado, toda vez que frente a entradas (inputs) isomorfas producen salidas (outputs) isomorfas. Esto, sin embargo, está completamente reñido con nuestra intuición.

Si la Teoría de la Identidad pecaba de excluir demasiados “organismos” del selecto grupo de aquéllos que poseen estados mentales, el Funcionalismo peca de incluir demasiados en el mismo grupo. Para Block el problema es un poco más grave. Según él, no hay forma de evitar esta paradoja. Si uno adscribe a la Teoría de la Identidad entonces se es demasiado chauvinista; si, en cambio, uno adscribe al Funcionalismo se es demasiado liberal, y no hay otra alternativa plausible. En otras palabras, no hay modo de caracterizar las entradas (inputs) y las salidas (outputs) de modo de evitar el chauvinismo y liberalismo simultáneamente (Block 1978).

Sin embargo, Sober observó que una versión teleológica del Funcionalismo permitiría evitar los anteriores inconvenientes (Sober 1985). Su postura se basa en la distinción entre el Funcionalismo “Máquina de Turing” (asociado a una noción matemática de función) y el Funcionalismo Teleológico. El problema que debe enfrentar Sober, como cualquier visión teleológica del Funcionalismo, es que no dispone de una noción clara, naturalizada de la noción de teleología. No ha habido una forma de identificar un objetivo último, que de sentido a los objetivos particulares que subsume, ni mucho menos hacerlo de un modo compatible con una concepción materialista del mundo. Y la tentación de apelar a Dios parece demasiado cerca...

Pues bien, como hemos visto a lo largo del presente texto, la Teoría de la Autopoiesis puede rescatarnos de esta terrible disyuntiva. Con ella podemos apelar a una versión teleológica del Funcionalismo sin caer en ninguna forma de deísmo o algo parecido. El Funcionalismo “Máquina de Turing” resulta, efectivamente, ser demasiado liberal, pero no porque la

caracterización de sus entradas y salidas sea poco específica, sino porque la descripción del proceso que una secuencia causal de estados mentales constituye es demasiado acotada, y no permite comprender el rol que ésta juega en un proceso mucho mayor que lo contiene, a saber, la autopoiesis. En el caso del grupo de personas contratadas para emular la actividad neuronal de un cierto estado mental, tendríamos que, simplemente, dicho grupo no participa de proceso autopoietico alguno y, por lo tanto, no pueden constituir un estado mental, es decir, no es psicopoietico. Un subproceso autopoietico (e.g. una configuración neuronal) es un mero proceso físico considerado separado del proceso completo de la autopoiesis.

Ni aunque, como también ha propuesto Block, sacásemos el cerebro de una persona y conectáramos todos los terminales de entradas y salidas del cuerpo del modo adecuado con las entradas y salidas del grupo de personas contratadas para emular el trabajo de neuronal del cerebro, realizando éstos los mismos procesos que aquél realiza; aun así, dudo que el sistema integrado sería autopoietico. No creo que dicho sistema soportaría el más leve análisis contrafáctico (Baste considerar el hecho que ofreciéndoles a dichas personas una suma de dinero mayor a lo que se les ofrece por golpear con el martillo a sus compañeros, presumiblemente dejarían de hacerlo, rompiendo con ello la autopoiesis del sistema.).

No hay actividad mental en sistemas que no sean vivos; esto es perfectamente intuitivo. Consecuentemente, no hay actividad mental sin antes haber autopoiesis. No sólo los estados mentales han de ser individuados funcionalmente sino que la misma mente también ha de serlo. Sin participar del proceso autopoietico no hay mente. La autopoiesis es una condición *sine qua non* para la psicopoiesis y, en consecuencia, para la actividad mental; ella es identificable con el objetivo final requerido por una versión teleológica del Funcionalismo.

## Capítulo 5

El paso de los años no logró consolidar del todo las expectativas que la aparición de las doctrinas funcionalistas de la mente trajo consigo. Sin lugar a dudas, el aspecto más flagrante de este fracaso es el problema del contenido intencional. La semántica del rol funcional, respuesta natural de dichas doctrinas a este viejo problema, encontró una serie de inconvenientes teóricos que le restan credibilidad (Fodor y Lepore 1992). Más grave aun, las respuestas alternativas aparecidas en los últimos 30 años, desarrolladas en el marco de una concepción funcionalista, tampoco han estado exentas de cuestionamientos teóricos insolubles (ver capítulo 2). Cabe, entonces, realizar un breve análisis suficientemente general de las razones que han socavado los intentos de reducir este problema.

En primer lugar, consignemos que el problema en cuestión es de orden metafísico. Existe algo que llamamos el contenido intencional, propio de todo estado mental intencional, que se expresa en una relación entre dicho estado mental y una porción de mundo. Esta relación es la que cotidianamente llamamos “ser acerca de”. Así, por ejemplo, toda creencia (un posible estado mental) instanciada en algún sujeto cognoscente participa de esta relación en tanto que es creencia *acerca de* algo. Pues bien, si estamos dispuestos a aceptar que la propiedad de nuestras creencias, deseos, miedos u otros estados mentales intencionales de ser *acerca de* algo no es meramente ilusoria, entonces debemos asignar al problema de develar dicha relación el carácter de metafísico (ver capítulo 2). Una solución del problema, por lo tanto, debe ser capaz de reducir la relación “ser acerca de” a los fundamentos ontológicos considerados aceptables dentro del marco en que se lleva a cabo la propuesta, cualesquiera que éstos sean. En el caso de la tradición de la filosofía analítica, en el cual se enmarca el presente trabajo, existe consenso en que estos fundamentos son los del materialismo.

En segundo lugar, las propiedades metafísicas que pueden servir de sustrato para la explicación de la intencionalidad pueden clasificarse en base a un esquema estímulo/mente/respuesta, o bien, para evitar el desprestigio filosófico asociado a palabras como estímulo y respuesta, percepción/pensamiento/conducta. La actividad mental, en tanto que ubicada entre la percepción (estímulo) y la conducta (respuesta), ofrece tres alternativas para explicar su extraña conexión con el mundo, llamada, ya está dicho, la propiedad de “ser

acerca de” porciones de éste. La primera consiste en intentar esclarecer dicha conexión recurriendo a fenómenos ubicados en el dominio de la percepción. Así, tomar esta alternativa supondría que la propiedad de “ser acerca de” de un estado mental se debe a fenómenos temporalmente previos a la instanciación del estado mental. Ejemplos de este recurso son las llamadas teorías causales (Stampe 1977) y otras derivadas (Fodor 1987, Dretske 1981). Una segunda alternativa sería establecer que la relación semántica ha de ser explicada sólo en términos de la actividad mental, prescindiendo de los dominios anteriores y posteriores a ella como la percepción y la conducta. En esta visión, a diferencia de la anterior, se supone que la propiedad de los estados mentales de “ser acerca de” es inherente a la actividad mental en sí misma. Esta es la categoría en que se ubican las semánticas del rol funcional denominadas “de un factor”. Por último, queda la opción de acudir a fenómenos en el dominio de la conducta para explicar la conexión entre estados mentales y mundo. La relación “ser acerca de” quedaría, entonces, determinada por fenómenos temporalmente posteriores a la instanciación de un estado mental. Si bien esta opción a sido la menos explorada de las tres, podríamos considerar dentro de esta categoría a algunas versiones de las semánticas del rol funcional “de dos factores” (Van Gulick 1980).

La gran ventaja de realizar esta taxonomía de las teorías en oferta para la relación entre mente y mundo es que permite vislumbrar la dificultad que éstas enfrentan en términos de dos intuiciones mutuamente excluyentes. Por un lado, la intuición de que la propiedad de los estados mentales de ser *acerca de* una porción de mundo debe ser inherente a los mismos aboga por la segunda alternativa (semánticas del rol funcional “de un factor”) en desmedro de las otras dos. Por otro lado, la intuición de que la relación en cuestión no es ilusoria y que tiene, pues, existencia y relevancia ontológica nos fuerza a exigir a cualquier explicación de ella un soporte metafísico con el mundo; bajo peno de renunciar al monismo ontológico o incurrir en el solipsismo. Este soporte es precisamente de lo que carecen las semánticas del rol funcional y lo que sí ofrecen las otras dos alternativas.

Renunciar a la inherencia del contenido intencional a los estados mentales conduce a inconvenientes, según creo, irreparables. Ejemplo de estos son las semánticas causales, las cuales pasan por alto que la historia –causal u otra- carece de potencia explicativa metafísica, por pertenecer al dominio epistémico (ver capítulo 2). Renunciar a la percepción y a la acción, únicas conexiones metafísicas entre mente y mundo, nos conduce inexorablemente al solipsismo; abnegando, con ello, del materialismo ontológico y devolviéndonos a una

concepción del problema Mente/Cuerpo en primera persona, propia de la filosofía moderna (Lycan 1990).

La única conclusión posible de esta encrucijada es que algo está fallando en los cimientos mismos de nuestra teoría de la mente. El capítulo anterior fue un llamado de atención respecto de este mismo hecho y un intento de postular una nueva noción de función para resolver el problema; a saber, la noción teleológica de función psicopoiética. Indagar en la aplicación de la idea de psicopoesis al problema del contenido intencional es el propósito del presente capítulo.

## **Contenido Conceptual y Función Psicopoiética**

Una teoría completa del contenido intencional requeriría ocuparse de un sinnúmero de detalles. Desde el problema del contenido conceptual hasta el problema de la referencia de los enunciados informativos (¿una porción de mundo o sus condiciones de verdad?) (Frege 1892) pasando por el de la categorización, el de la confirmación, etc. Una empresa así difícilmente pueda ser llevada a cabo por un solo hombre en un solo esfuerzo filosófico. Será, pues, necesario acotar el problema. Intentaremos, por lo tanto, entregar, si bien no una respuesta concreta a todos los detalles involucrados, al menos un marco general desde donde construirlas y, por sobre todo, una explicación plausible que muestre la factibilidad de explicar metafísicamente el contenido intencional al modo naturalista, compatible con el materialismo ontológico.

La primera hipótesis, más o menos habitual, que nos ayudara a dar forma al problema es que el contenido intencional de las representaciones mentales (i.e. el contenido de las creencias, deseos, etc.) se construye, de algún modo, del contenido de los conceptos (primitivos) que participan de ellas, partículas básicas de la actividad mental. A pesar de contar con algunos detractores (Loar 1983), esta hipótesis es más que aceptable y hasta necesaria para quienes adhieren a las teorías representacionales de la mente (Fodor 1998).

En segundo lugar, obviando ciertas distinciones útiles en ciertos contextos de análisis (Harman 1987) no haremos mayor diferencia entre contenido y significado conceptual. Más generalmente, nos ocuparemos del problema de esclarecer, sobre la base de una ontología materialista, cuáles son los fundamentos metafísicos que subyacen tras la conexión entre

conceptos y objetos del mundo, cotidianamente referida como la propiedad de los primeros de ser *acerca de* los segundos.

En tercer lugar, supondremos que la propiedad de ser *acerca de* algo es exclusiva de las entidades mentales. Más de alguno no estaría de acuerdo con esta sentencia (Millikan 1984, Dretske 1981); sin embargo, la propuesta es aceptable si se considera que la intencionalidad es, en una concepción materialista (y, por lo tanto, monista), aquello que permite caracterizar la actividad mental como fenómeno peculiar distinto de otros fenómenos físicos (ver capítulo 1). Quienes han optado por trivializar la intencionalidad extendiéndola a toda clase de entidades como forma de naturalizarla, olvidan que una de las principales razones para querer naturalizarla es avanzar en la resolución del problema Mente/Cuerpo. Sin embargo, trivializar la intencionalidad supone renunciar a la mejor caracterización que tenemos de lo mental y, por lo tanto, constituye un retroceso en el sentido deseado.

Pero si esto no bastara, podemos decir que desde una perspectiva intuitiva es difícil concebir la intencionalidad como fenómeno no-mental. En efecto, en ausencia de sujetos cognoscentes alrededor (e.g. antes de que existiera la vida en el universo), difícilmente podría un símbolo físico (e.g. una imagen, un sonido, etc.) ser *acerca de* algo. Asumiremos, por el contrario, que la aparente capacidad de los símbolos físicos de ser *acerca de* objetos, propiedades o situaciones se debe a la tendencia convencional de los sujetos portadores de mente que los rodean a instanciar conceptos o símbolos mentales que, de hecho, son *acerca de* esos objetos, propiedades o situaciones, cuando están en presencia de aquellos.

Hechas estas salvedades, podemos entrar en materia. Según dijimos, el gran obstáculo que las semánticas del rol funcional “de un factor” no lograban sortear era el solipsismo. Intentar explicar la propiedad de los conceptos *de ser acerca de* algo en términos de fenómenos de alcance restringido al solo ámbito mental deja a estas teorías sin una conexión metafísica con el mundo; cuestión que es inadmisiblesi lo que se quiere explicar es, precisamente, un tipo de conexión entre entidades mentales y mundo. Por mi parte, postularé que estas propuestas no están completamente erradas y que su compromiso con el solipsismo se debe a una noción de función errónea. Más aun, creo que el contenido conceptual está estrechamente vinculado con el rol funcional de éstos en la actividad mental, pero bajo una concepción de función teleológica y no Máquina de Turing. Específicamente, postularé como candidato a dicha noción la de función psicopoiética, desarrollada en el capítulo anterior.

Función psicopoiética es una noción teleológica, pues, según dijimos, su descripción no está restringida a la noción matemática de función, es decir, como mapeo de estados mentales en otros estados mentales. Más bien, la función psicopoiética de un concepto ha de entenderse como el aporte metafísico de dicho concepto a un proceso global, denominado autopoiesis del organismo en cuestión. Ahora bien, si la autopoiesis de los organismos es, justamente, lo que los caracteriza como objetos vivos en términos estrictamente materialistas, la noción de psicopoesis, un tipo de proceso autopoietico especializado en la relación con el entorno, exhibe la conexión metafísica con el mundo que la noción de función requiere para dar sentido a una semántica funcionalista. Más aun, lo hace sin sacrificar la intuición de que la propiedad de los conceptos de ser *acerca de* objetos o propiedades es intrínseca a los mismos, cuestión que pocas semánticas son capaces de garantizar.

Conforme a lo establecido en el capítulo anterior, los procesos psicopoiéticos tienen la propiedad de ser específicos respecto de objetos del mundo. Así, por ejemplo, el proceso psicopoiético producido por el deseo de beber cerveza en Clyde (ver capítulo 4) produce cambios en la relación entre el sujeto, Clyde, y el objeto, cerveza. Esta característica resulta fundamental para explicar la propiedad del concepto CERVEZA de ser *acerca de* cervezas. En efecto, en lo que resta de este capítulo y en el próximo, defenderemos la tesis de que un concepto  $X$  es *acerca de* un objeto  $x$  en la medida en que los procesos psicopoiéticos a que dan lugar las actitudes proposicionales en que participa  $X$  son específicos respecto de  $x$ . En otras palabras, la propiedad de los conceptos de ser *acerca de* objetos corresponde a la capacidad de aquéllos de aportar *especificidad* a los procesos psicopoiéticos producidos por las actitudes proposicionales en que participa.

Con todo, como acertadamente observa Fodor (1987), cualquier semántica filosófica que se precie de tener alguna validez, debe ser capaz de explicar la relación entre símbolo y objeto sin contener dentro de su lenguaje referencias a propiedades semánticas, bajo pena de caer en una “petición de principio”. En el caso que aquí nos ocupa, el principal riesgo de incurrir en esta falta se encuentra en la noción de especificidad. En ella se concentra toda la carga teórica de nuestra propuesta y deberemos, por lo tanto, ser capaces de explicarla en términos que excluyan cualquier referencia a propiedades semánticas. Esta será nuestra siguiente tarea.

## Especificidad

En el capítulo anterior planteamos que un proceso psicopoiético era específico respecto de un objeto en la medida en que aquél cambiaba la relación entre el sujeto en que se instanciaba el proceso y el objeto en cuestión. No obstante lo tentadora que esta explicación pueda sonar, su sencillez augura dificultades. Acaso la mayor de ellas sea la imposibilidad de distinguir objetos del mundo que esta explicación supone. En efecto, la forma natural de definir cual es la “relación” entre sujeto y mundo es por medio de la disposición espacial entre éstos. Sin embargo, cualquier movimiento que un organismo realice supone cambios en la disposición espacial respecto de todos los objetos del mundo. Si, por ejemplo, Clyde se encuentra sentado en un sillón viendo un partido de fútbol y, de pronto, siente deseos de beber cerveza y acerca su mano a la lata de cerveza que se encuentra sobre la mesa, este movimiento cambia su disposición espacial respecto de la lata de cerveza, pero también respecto de la mesa, del sillón y del televisor. ¿Cómo podremos decir, entonces, respecto de qué objeto es específico el proceso psicopoiético a que dio lugar su deseo? La relación espacial no será la solución.

Pues bien, dejando a un lado los rodeos, existe una forma de definir especificidad de modo completamente naturalista. En efecto, diremos que un proceso psicopoiético es *específico* respecto de un objeto o entidad del mundo cuando el éxito o fracaso de dicho proceso dependa necesariamente (en el sentido modal del adverbio) de las propiedades metafísicas de dicho objeto o entidad. En otras palabras, cuando el proceso, caracterizado por las actitudes proposicionales que lo produjeron, sea exitoso o no en la conservación de la regularidad autopoiética dependiendo, en todo mundo metafísicamente posible, de las propiedades del objeto o entidad en cuestión. Esto no quiere decir que el proceso deba ser exitoso para que sea específico de un cierto objeto. Por el contrario, proponemos que será específico si, tenga éxito o no, esto depende necesariamente de las propiedades metafísicas del objeto.

El adverbio modal “necesariamente” resulta fundamental para que la definición de especificidad pueda satisfacer las exigencias. La razón es que, de no mediar dicho adverbio, podría darse el caso que el éxito o fracaso del proceso psicopoiético dependiera contingentemente de propiedades de otros objetos distintos del relevante. Por ejemplo, en el caso del capítulo anterior en que Clyde desea beber cerveza y cree que hay una lata en el refrigerador, la presencia del perro en su camino hasta la cocina hace que el éxito del proceso psicopoiético a que dieron lugar el deseo y la creencia *acerca de* cerveza dependa también del objeto “perro”. Consecuentemente, nuestra definición de especificidad no lograría establecer

por qué el proceso psicopoiético descrito es específico respecto de cerveza y no de perro. Sin embargo, resulta obvio que el hecho de que el perro esté en el camino de Clyde a la cocina es contingente, como también la dependencia del proceso de las propiedades del objeto “perro”. En oposición, la satisfacción del deseo de Clyde de beber cerveza depende, en todo mundo imaginable, del objeto “cerveza”, así como el éxito o fracaso del proceso psicopoiético, caracterizado por dichas actitudes, de las propiedades del mismo.

Por otro lado, las creencias se conciben usualmente como susceptibles de ser evaluadas en términos de verdaderas o falsas y los deseos en términos de satisfechos o insatisfechos, nociones habitualmente consideradas semánticas. La similitud de estas nociones evaluativas con las de éxito y fracaso bien podría levantar dudas respecto de la ausencia de connotaciones semánticas en nuestra definición de especificidad. En mi defensa debo decir que, más allá de las discusiones respecto del real carácter semántico de las nociones de satisfacción o insatisfacción y verdadero o falso (Harman 1987) la mejor garantía de ausencia de connotaciones semánticas en una descripción es la posibilidad de traducirla a lenguaje estrictamente fisicalista. Y en nuestro caso, afortunadamente, es posible realizar este ejercicio. En efecto, el éxito o fracaso de un proceso psicopoiético puede traducirse como la conservación o no conservación de un conjunto de regularidades físicas, denominadas autopoiesis, (ver capítulo 4) que distinguen la unidad espacial en que éstas tienen lugar de los objetos cotidianamente entendidos como inertes. Si esto tiene connotaciones semánticas también las tiene el Principio físico de Conservación de la Energía.

Una gran virtud de esta definición de especificidad de los procesos mentales intencionales, aquí referidos como psicopoiéticos, es que respeta una intuición que, hasta donde yo sé, ninguna otra semántica consigue rescatar. Ésta es que sólo entes vivos pueden tener actividad mental intencional. Bajo una semántica del rol funcional, construida sobre una noción de función “matemática”, todo ente, vivo o no, que exhiba procesamiento de información al estilo Máquina de Turing (e.g. un computador), posee estados con contenido intencional, en tanto que éstos poseen una función (i.e. realizan mapeo a otros estados). Bajo una semántica causal, cualquier ente que contenga símbolos causados por objeto o propiedades del exterior es portador de estados intencionales (i.e. *acerca del* objeto en cuestión). La consecuencia previsible de esta falencia es que todas estas semánticas se ven obligadas a restringir independientemente lo mental al ámbito de lo vivo sobre la base de características fisiológicas generalmente; o, simplemente, deben defender concepciones liberales (en el sentido de Block 1978) respecto de la mente. Yo, por mi parte, no me debo a ninguna de estas obligaciones. El

hecho de que la propiedad “ser acerca de” remita a la conservación o no conservación de regularidades características de los entes vivos –como es la autopoiesis- excluye cualquier posibilidad de que entes inertes exhiban estados mentales intencionales.

Más allá de esto, individuaremos los conceptos funcionalmente. Esto es, como entidades cuya función es hacer que ciertos procesos (i.e. los psicopoiéticos) característicos de ciertos organismos vivos (i.e. los dotados de sistema nervioso central) dependan necesariamente de las propiedades metafísicas de un(os) cierto(s) objeto(s). En otras, palabras, los conceptos serán entidades capaces de orientar la dependencia de los procesos psicopoiéticos a objetos del mundo. Más aun, identificaremos el contenido de dichos conceptos con el cúmulo de propiedades metafísicas asociadas al objeto y no con el objeto mismo. Con esto podremos independizar el contenido conceptual de la existencia del objeto *acerca del* cual es el concepto; como veremos más adelante, cuando analicemos el problema de las Tierras Gemelas en relación con nuestra teoría.

A esta forma de explicar la noción de intencionalidad en términos de una más dócil – metafísicamente hablando- de dependencia llamaremos Semántica de la función psicopoiética. A continuación someteremos esta visión del contenido conceptual a las objeciones más habituales en contra de las semánticas del rol conceptual. Antes, sin embargo, resulta de vital importancia realizar dos advertencias concernientes a la presente teoría.

En primer lugar, que al identificar el contenido conceptual con un cúmulo de propiedades metafísicas asociadas habitualmente (aun cuando no necesariamente: ver, más abajo, discusión respecto de las Tierras Gemelas) a un objeto del mundo no estamos suponiendo el conocimiento por parte del sujeto poseedor del concepto de dichas propiedades. Muy por el contrario, como veremos cuando discutamos el problema del holismo semántico, la red de creencias en que un concepto participa no determina de ningún modo la función psicopoiética del mismo. Como resultado tendremos que los procesos inferenciales (también llamados computaciones) respetan el contenido de los conceptos que participan de ellos, cuestión considerada por algunos imprescindible para una teoría de la mente (Fodor 1998).

En segundo lugar, que esta explicación del contenido conceptual no intenta abarcar las razones que hacen que existan objetos mentales que tengan la función de aportar especificidad a los procesos psicopoiéticos característicos de los seres vivos dotados de mente. Ésta no es una teoría de la formación de los maravillosos y extraños objetos que son los conceptos, sino

que, supuesta su existencia, de la relación metafísica de éstos con porciones de mundo. Con todo, es plausible pensar en ellos como un rasgo filogenético, desarrollado a partir de una primitiva capacidad de los seres vivos de responder selectivamente a las características de su entorno. Esta capacidad se observa incluso en organismos carentes de sistema nervioso central. Las plantas, por ejemplo, exhiben mayor propensión al crecimiento en direcciones abundantes en luz solar. Acaso éste podría ser un indicio muy primitivo de la capacidad de los organismos dotados de mente, ubicados en un escaño evolutivo superior, de tener procesos psicopoiéticos dependientes de las propiedades de los objetos del mundo (e.g. de sol); o, en otras palabras, de tener actitudes proposicionales *acerca de* algo.

## **Error**

Un problema habitual para casi todas las semánticas filosóficas es el llamado problema del error. De modo general, podemos decir que la dificultad que plantea es que los sujetos cognoscentes de tanto en tanto cometen errores al aplicar conceptos a objetos del mundo. Por ejemplo, mucha gente en presencia de coyotes puede equivocarse al instanciar estados mentales *acerca de* perros. Pues bien, una teoría en que el contenido de dichos estados depende de los objetos externos conectados metafísicamente con ellos por medio de la percepción o la conducta, corre el riesgo de no considerar la posibilidad de error. En nuestro ejemplo de la persona que confunde coyotes con perros, una teoría tal podría indicar que el estado mental que se instancia en presencia del coyote es *acerca de* coyotes. Sin embargo, sabemos, por construcción del ejemplo, que el estado mental instanciado es *acerca de* perros. Luego, la teoría falla en su predicción.

Las semánticas del rol funcional no son ajenas a esta dificultad a pesar de obviar tanto el dominio de la percepción como el de la conducta en su explicación del contenido. La razón es que si, como postulan, el contenido intencional depende de la red de inferencias en que los estados mentales participan, entonces, cabe la pregunta de si las inferencias erróneas son también constitutivas del contenido. Lamentablemente estas semánticas no cuentan con una forma no-intencional plausible de distinguir las inferencias correctas de las erróneas. Esta crítica sigue siendo muy dañina para ellas.

En el caso de nuestra semántica de la función psicopoiética, el problema del error también puede resultar inconveniente. Para ilustrar las dificultades, consideremos, en primer lugar, el mismo caso de una persona que, parada frente a un coyote, lo confunde e instancia un estado

mental *acerca de* perro. La pregunta fundamental es ¿acerca de qué, indica nuestra teoría, será el estado mental instanciado? La respuesta puede encontrarse traduciendo la pregunta al lenguaje de nuestra teoría. ¿Cuál es la función psicopoiética del concepto que participa del estado mental, aportar especificidad respecto de perro o de coyote?

Para intentar dar respuesta a la pregunta, recordemos, en primer lugar, la definición de función psicopoiética. Según dijimos, la función psicopoiética de un concepto corresponde a la determinación de las propiedades metafísicas que harán necesariamente fracasar o tener éxito a *todos los procesos psicopoiéticos* producidos por las actitudes proposicionales en que aquél participa. A primera vista, uno tiene la impresión de que cualquier proceso psicopoiético a que dé lugar el concepto instanciado en la mente de la persona que confunde un coyote con un perro dependerá del coyote, que es el que, de hecho, está situado en frente de dicha persona. Este hecho implicaría que nuestra teoría predice que el concepto instanciado es *acerca de* coyotes y no *acerca de* perro; cuestión, por cierto, indeseable. Sin embargo, esto no es rigurosamente correcto. Más preciso sería decir que cualquier proceso psicopoiético a que dé lugar el concepto instanciado en la mente del sujeto que comete el error dependerá del objeto indeterminado –sea éste perro, coyote, caballo, mesa, etc.- que está enfrente de él. Esto resulta evidente al considerar que en muchos procesos motivados por actitudes proposicionales en que participa el concepto PERRO también participa el objeto (indeterminado) que está situado en frente del sujeto. Por ejemplo, podríamos aceptar que en la mente de nuestro individuo ocurre la siguiente secuencia de actitudes proposicionales: (1) El deseo de acariciar a un PERRO; (2) La creencia de que ESO (el coyote, que está frente a él) es un PERRO. Consecuentemente el proceso psicopoiético caracterizado por estas actitudes (i.e. acercarse y acariciar a “eso”) dependerá necesariamente del coyote que, de hecho, está frente a él; pero esto será gracias a que hay un concepto (ESO) que aporta especificidad respecto de ese objeto en las actitudes, y no gracias al concepto PERRO. Este último seguirá aportando especificidad respecto de perros y por lo tanto seguirá siendo *acerca de* éstos.

En el fondo de este sutil argumento está un hecho que las semánticas causales, la biosemántica y muchas otras simplemente no pueden respetar; esto es, que en una descripción metafísica de un mundo (i.e. una que no apele a capacidades cognitivas) no hay error. En otras palabras, el error sólo aparece en el mundo cuando aparecen los sujetos cognoscentes (*a fortiori*, cuando aparecen los hombres). Sin embargo, lo que hemos tratado de argumentar a través de toda esta investigación es que la intencionalidad es una propiedad metafísica

presente en el mundo ontológicamente previa a las capacidades cognitivas de los sujetos. Consecuentemente, la intencionalidad debiera figurar en una descripción metafísica del mundo y, por lo tanto, debiera estar libre de error. Luego, no puede haber error semántico asociado a las entidades mentales. No puede ser el caso que un concepto signifique erróneamente algo. Si es que hay error este debe estar localizado en la estructura cognitiva. Si hay error este sólo puede encontrarse en nuestro acceso epistémico al mundo y por ningún motivo en nuestro acceso metafísico.

Conforme con lo anterior, nuestra forma de encarar el problema del error consiste simplemente en separar la categorización, cuestión ciertamente epistémica, de la semántica conceptual, cuestión metafísica. Yo puedo instanciar el concepto PERRO en presencia de un coyote, pero este concepto sigue siendo acerca de perros: no hay error semántico. El error aparece cuando instancio en mi mente (2) la creencia de que ESO (que está en frente) es un PERRO. Aquello que describimos como ESO refiere al coyote que está frente a mí; PERRO refiere a perros y el error está en creer que lo referido por el primero pertenece a la clase referida por el segundo. Esto, sin embargo, no va más lejos de la realización de un juicio equívoco respecto del mundo, cuestión ciertamente epistémica. ESO sigue significando el objeto situado frente a mí y PERRO sigue significando perro.

Por otro lado, note el lector que hemos destacado la frase “todos los procesos psicopoiéticos” en la definición de función psicopoiética anterior. Esto para indicar que no es posible, en nuestra teoría, evaluar cuál es el contenido de un concepto en virtud de una sola instanciación del mismo. En estricto rigor, desde la perspectiva del observador externo, ningún número finito de casos podría determinar certeramente el contenido de un concepto. Sin embargo, negando la definición anterior, sí podemos asegurar cuando un concepto no es *acerca de* algo. Bastaría encontrar un proceso psicopoiético cuyo éxito o fracaso no dependa necesariamente de ese “algo”. En nuestro caso, para demostrar definitivamente que el concepto PERRO sigue significando perro bastará con imaginar que la presencia del coyote confundido con perro por el observador, trae consigo una serie de actitudes proposicionales en cuya descripción jamás aparece mencionado el objeto que está en frente (i.e. no figura un “ESO”). Construyamos un ejemplo tal.

Como primer paso, convengamos que, si es que los perros son metafísicamente distintos de los coyotes, éstos deben diferir en alguna propiedad. Una propiedad más que aceptable puede ser la hostilidad que los coyotes profesan a los humanos versus la amistad de los perros hacia

los mismos. Supongamos, entonces, que nuestro protagonista es un transeúnte que, camino a casa, divisa en la lejanía un coyote y lo confunde con un perro. Supongamos, a continuación, que en su mente se instancian las siguientes actitudes proposicionales: (1) El deseo de jugar con un PERRO; (2) La creencia de que su perro, BOBY, está esperándolo en casa. (Notar la ausencia del concepto ESO, referido al coyote.) Consecuentemente, el proceso psicopoiético caracterizado por éstas actitudes se traducirá en que nuestro protagonista apure el paso a su casa. Naturalmente, el éxito o fracaso de dicho proceso no dependerá necesariamente de las propiedades del coyote –sino de “perro” y “Boby”- aun cuando las actitudes que lo produjeron pudieron haber sido motivadas por el coyote. Nuestra teoría, consecuentemente, predice que el concepto instanciado en la mente del sujeto no será *acerca de* coyotes sino *acerca de* perro; eludiendo, con ello, el problema del error.

## **Propiedades Sensoriales**

Otra crítica en contra de las semánticas del rol conceptual fue planteada por Fodor a propósito de las formas de percepción sensorial. Según él, las inferencias que se realizan en la mente de un sujeto cognoscente involucran, tarde o temprano, a la evidencia sensorial. Así, bajo una semántica del rol conceptual, dos personas con distinta capacidad de percepción, a la vez que poseen distinta evidencia sensorial, poseerían también distintos conceptos *de* un mismo objeto. Fodor propone el ejemplo de Helen Keller, quien es ciega y sorda de nacimiento, para visualizar el problema. Si el contenido conceptual está determinado por (ni mencionar si es idéntico a) las relaciones inferenciales en que el concepto participa, entonces, en la medida en que Helen Keller porta distinta evidencia sensorial que yo, las inferencias en que participan su concepto de agua y el mío son distintas. En consecuencia, nuestros conceptos de agua tienen distinto contenido, cuestión que, obviamente, atenta contra la publicidad de los conceptos (Fodor 1998).

El efecto esperable es que Helen Keller y yo no hablamos de lo mismo cuando hablamos de agua. Sin embargo, es evidente que ella y yo podríamos tener una conversación *acerca de* agua. ¿Cómo puede ser esto posible si nuestros conceptos significan cosas distintas?

Para nuestra teoría, en cambio, la percepción sensorial no significa el más leve problema gracias a que no considera las relaciones inferenciales como constitutivas del contenido. Por el contrario, aun cuando las inferencias que se instancian en un sujeto cognoscente en ocasiones pueden considerarse parte de uno o varios procesos psicopoiéticos (como resulta

evidente cuando uno esta haciendo filosofía u otra actividad intelectual) y, aun cuando, el éxito de un proceso psicopoiético dependerá frecuentemente de realizar “buenas” inferencias, la función psicopoiética de los conceptos que en ellas participen no se verá alterada por cuáles inferencias se realicen. En este punto coincidimos con Fodor (1998), quien postula que los procesos inferenciales (también llamados computaciones) respetan el contenido conceptual.

Volveremos sobre este asunto cuando nos defendamos de posibles imputaciones de holismo semántico. De momento convengamos que si Helen Keller tiene actitudes proposicionales en que participa el concepto AGUA, a pesar de que la información sensorial que ella posee respecto de agua es limitada, lo cual podría llevarla a tomar malas decisiones, el éxito o fracaso de los procesos psicopoiéticos producidos por dichas actitudes seguirá dependiendo necesariamente de las propiedades del agua. Consecuentemente, la función psicopoiética del concepto es aportar especificidad respecto de agua, al igual que la del mío o de cualquier otro individuo que lo posee. El concepto de Helen y el mío tienen, pues, el mismo contenido y no es extraño que podamos conversar *acerca de* agua, ni tampoco lo es que podamos comprendernos.

## **Tierras Gemelas**

Putnam (1975), el propio precursor de las teorías funcionalistas de la mente, arrojó un problema que atañe a casi todas las teorías del contenido. Se trata de un problema bastante conocido, de modo que iremos rápido en su planteamiento. Supongamos que en un lugar remoto del universo existe una tierra gemela que es idéntica a la nuestra en todo aspecto excepto en una leve diferencia en la composición molecular del agua; siendo en ella XYZ en lugar de H<sub>2</sub>O. Al ser idéntica la tierra gemela en todo otro aspecto, todos debemos tener un gemelo idéntico allá. Pues bien, el problema es que, a pesar de que cuando mi gemelo y yo tenemos pensamientos simultáneos acerca de agua estamos presumiblemente en el mismo estado psicológico, el sentido común indica que nuestros conceptos de agua poseen distinto contenido, pues, en estricto rigor, son *acerca de* objetos distintos; a saber, de H<sub>2</sub>O y XYZ, respectivamente.

Esta objeción resulta especialmente severa con las semánticas del rol conceptual de “un factor”. Al no disponer éstas de una conexión metafísica entre conceptos y mundo exterior no son capaces de dar una explicación a la intuición de que los conceptos de agua de mi gemelo y mío tienen distinto contenido.

Con todo, la necesidad de salvaguardar la inherencia del contenido a los conceptos (más generalmente, a los estados mentales) ha llevado a proponer una noción de contenido de composición mixta. Esto se expresaría en dos tipos de contenido, llamados contenido amplio y contenido estrecho. El primero apoyaría la idea de que éste, de un modo u otro, depende fuertemente de la extensión del estado mental. Consecuentemente, el contenido amplio de una creencia o deseo sería susceptible de evaluación semántica; esto es, en términos de verdadero o falso. El segundo, en cambio, aportaría la intuición de que el contenido debe depender de propiedades inherentes al sujeto. El contenido estrecho sería, pues, equiparable a lo que desde Frege se ha llamado la intensión de un estado mental y no admitiría evaluación en términos de verdadero o falso, sino que dependería estrictamente de la estructura psicológica de la mente en que se instancia.

Según esta definición dual del contenido los pensamientos *acerca de* agua de mi símil en la tierra gemela y míos difieren en contenido amplio y coinciden en contenido estrecho. Sin embargo, este recurso de apariencia arbitraria concebido para compatibilizar dos intuiciones mutuamente opuestas respecto del contenido no ha concitado demasiada adhesión. Naturalmente, la noción de contenido amplio ha sido generalmente eximida de cuestionamientos sustantivos situándose toda la discusión en torno a la existencia del contenido estrecho.

La abundancia de argumentos en contra de esta última noción de contenido (Putnam 1975, Burge 1979, Fodor 1987, etc.) no ha bastado, sin embargo, para acabarla. La razón fundamental está, creo, en la ya mencionada intuición de que el contenido debe ser inherente a los estados mentales o conceptos, junto con la ausencia de una teoría capaz de compatibilizarla con posiciones externalistas. La principal dificultad radica en los casos de conceptos carentes de extensión; como son, por ejemplo, los conceptos de bruja, flogisto, u otros similares. La pregunta que los partidarios del contenido estrecho han esgrimido como argumento es ¿cómo es posible que el contenido de ciertos conceptos dependa de objetos inexistentes? Un hecho en favor de la teoría que estamos exponiendo en el presente trabajo es que consigue eludir este inconveniente.

En efecto, deliberadamente omití en la definición de función psicopoiética la referencia directa a objetos, en favor de una más neutral a las propiedades de éstos. La idea era no comprometer la existencia de los mismos en la noción de contenido. Por el contrario, *stricto dicto*, lo único que se requiere en nuestra teoría para que un concepto sea *acerca de* algo es un

cúmulo de propiedades y no la existencia efectiva de ese “algo”. Innegable es el hecho que la existencia efectiva de un objeto define de manera inequívoca un cúmulo de propiedades, sin embargo, esto también puede hacerse por vías puramente convencionales. El caso de los conceptos de bruja o flogisto son claramente de este último tipo. Podríamos decir acaso que es admisible que objetos inexistentes tengan propiedades (e.g. en el caso de las brujas, andar en escobas), pero esto supondría una intromisión innecesaria en discusiones metafísicas respecto del carácter predicativo de la existencia. Preferible es conformarse con la intuición de que la ontología de objetos que reconocemos es, en último término, hipotética; junto con una definición de contenido conceptual que considera la posibilidad de que los objetos constitutivos del contenido no existan *de facto* sino tan sólo convencionalmente por medio de cúmulos de propiedades.

Lamentablemente, esta salida al problema de los conceptos sin extensión no funciona para otras semánticas externalistas. Por ejemplo, en el caso de las semánticas causales, al ser la relación causal de carácter metafísico, no admiten la posibilidad de que el objeto que causa un cierto concepto no exista. La replica inmediata a una postura así sería que nada que no exista puede causar metafísicamente algo que sí existe. Sólo los objetos existentes poseen poderes causales (ver capítulo 2).

En nuestra teoría, en cambio, al concebir a los organismos vivos como máquinas cerradas, su conexión metafísica con el mundo posee un carácter relacional. Esto es, los objetos no poseen realidad metafísica para las máquinas autopoieticas sino tan sólo su impacto sobre (o relación con) ellas. Esta idea favorece una ontología fundamental compuesta de propiedades para los entes vivos en desmedro de una compuesta de objetos. Éstos resultan, por el contrario, de un acto epistémico a partir de dicha ontología fundamental; es decir, los objetos sólo poseen realidad epistémica para las máquinas autopoieticas. No es rara, entonces, la capacidad de los organismos dotados de mente de asignar existencia hipotética a cúmulos de propiedades. Realizar este ejercicio es parte natural de lo que somos y de cómo conocemos.

Por último, el rumbo de la discusión nos llevó a olvidarnos temporalmente del problema de las tierras gemelas. Si nuestra teoría del contenido conceptual ya consiguió compatibilizar una concepción externalista con la intuición de que el contenido es, de algún modo, inherente a los conceptos mismos, por un lado, y eludir el problema de los conceptos carentes de extensión, por otro, dos metas que ninguna semántica conocida ha reunido simultáneamente; veremos

que el argumento de las tierras gemelas no significa siquiera un inconveniente para ella. Podremos, por lo tanto, prescindir completamente de la noción de contenido estrecho.

Para probar este hecho recordemos que, en nuestra semántica, un concepto es *acerca de* un objeto (o, talvez deberíamos decir, de un cúmulo de propiedades) si, y sólo si, el éxito o fracaso de *todos los procesos psicopoiéticos* producidos por las actitudes proposicionales en que aquél participa depende necesariamente de las propiedades de este último. En el caso del problema de las tierras gemelas, queremos que nuestra teoría respete la intuición de que mi concepto de agua y el de mi símil en la tierra gemela tienen distinto contenido, pues sus extensiones difieren (una es H<sub>2</sub>O; la otra XYZ). Luego, debemos mostrar que al menos existe un proceso psicopoiético producido por actitudes proposicionales que involucran mi concepto de agua cuyo éxito o fracaso no depende necesariamente del agua-gemela. Así habremos probado que el contenido de mi concepto AGUA no es agua-gemela. Por simetría, el contenido del de mi símil en la tierra gemela tampoco será agua.

Pues bien, tan sólo debemos pensar en una propiedad en la que las aguas de nuestra tierra y la gemela difieran e imaginar, a continuación, un proceso psicopoiético cuyo éxito dependa de dicha propiedad. En este caso, las dos aguas difieren en su composición molecular, luego esta propiedad nos servirá. Pensemos, entonces, que un buen día se me ocurre investigar la composición molecular del agua que tenemos en nuestra tierra. Nunca he sabido manejar microscopios de modo que es esperable que no me vaya muy bien en mi empresa. Una cuestión es clara, sin embargo: el éxito o fracaso de mi gestión, caracterizada por un deseo en que participa mi concepto AGUA, no dependerá de las propiedades que tenga el agua-gemela. Puede que concluya erróneamente que el agua de nuestra tierra tiene una composición H<sub>3</sub>O, pero este fracaso nada tiene que ver con que el agua-gemela tenga otra XYZ. Es decir, este proceso psicopoiético particular no depende de las propiedades del agua-gemela. Luego, los procesos psicopoiéticos en que participa mi concepto AGUA no dependen necesariamente de las propiedades del agua-gemela y, por lo tanto, mi concepto no es *acerca de* ella.

## **Composicionalidad**

Dos rasgos empíricos de la mente abogan por la necesidad de toda semántica filosófica y lingüística de respetar el principio de composicionalidad. Primero, la llamada productividad; esto es, la capacidad no acotada de la mente de imaginar pensamientos y conceptos tan complejos como se quiera. Segundo, la llamada sistematicidad; es decir, la extraña capacidad

de la mente de elaborar pensamientos distintos manteniendo la estructura sintáctica de éstos y combinando conceptos con distinto significado que el sujeto posea. Es difícil, por ejemplo, dudar que todo aquél que puede instanciar el pensamiento “Juan ama a María” puede también instanciar el pensamiento “María ama a Juan” (Fodor 1998), simplemente intercambiando los conceptos MARÍA y JUAN en la misma estructura sintáctica.

Por otro lado, más allá de las discusiones entre partidarios de las arquitecturas seriales y conexionistas, ya nadie cuestiona el hecho de que, sea lo que sea la mente (Máquina de Turing o Conexionista), ésta debe implementarse materialmente, de donde, su capacidad de memoria debe ser finita. La única forma (conocida, al menos) de explicar este hecho sin poner en cuestión la productividad de la mente es aceptar que la mayor parte de los conceptos y pensamientos realizables en las mentes han de ser composiciones de un conjunto finito de conceptos básicos. Naturalmente, la sintaxis y el contenido de estos conceptos o pensamientos compuestos se heredan de los de sus conceptos constituyentes. Esto es lo que se conoce como el principio de la composicionalidad.

La fuerza de este principio radica en que, al tiempo de compatibilizar la finitud de la mente con su productividad, explica también la sistematicidad de ésta. En efecto, si los pensamientos y conceptos complejos no son más que combinaciones de otros básicos, resulta, pues, natural que las mentes sean capaces de intercambiar en una estructura conceptos con poderes sintácticos similares, aunque con distinto contenido, produciendo así un pensamiento igualmente coherente.

Con todo, la naturalidad con que este principio explica cuestiones tan disímiles no ha invitado a su respeto unánime. Fodor y Lepore (1992) han acusado a las semánticas del rol conceptual de incurrir en su violación. La razón radica en el hecho de que, bajo esta visión del contenido conceptual, hay conceptos estructurados cuyo contenido no se derivaría del contenido de sus constituyentes. Ejemplo de esto, dicen Fodor y Lepore, es el concepto SERPIENTE ZUMBANTE. Si el contenido conceptual corresponde a la función de los conceptos en la estructura de inferencias en que éstos participan, entonces, de la sentencia “esto es una serpiente zumbante” se infiere la sentencia “esto es peligroso” (i.e. una serpiente cascabel), la cual, sin embargo, no se infiere de ninguna de las sentencias “esto es una serpiente” o “esto es zumbante”. Dicho de otro modo, las inferencias en que participa el concepto SERPIENTE ZUMBANTE no están determinadas por las inferencias en que participan SERPIENTE y

ZUMBANTE. En consecuencia, el contenido de SERPIENTE ZUMBANTE no se deriva del contenido de sus constituyentes, contraviniendo, con ello, el principio de composicionalidad.

Se ha dicho, en respuesta a lo anterior, que esta crítica sólo atañe a las semánticas del rol conceptual holistas; esto es, a aquéllas que consideran a todas las inferencias como constitutivas del contenido. Quienes, por el contrario, restringen las inferencias relevantes al contenido podrían admitir que entre las restantes inferencias en que participa un concepto compuesto hubiese algunas que no se siguiesen de inferencias en que participan sus constituyentes. Este sería el caso de SERPIENTE ZUMBANTE y la inferencia “esto es peligroso”.

Desgraciadamente, sin embargo, el mismo argumento que las defiende de esta crítica termina hundiendo a las semánticas del rol conceptual no-holistas, pues, a continuación, deben responder a la pregunta respecto de cuáles, si no todas, son esas inferencias tan particulares que son constitutivas del contenido y por qué razón éstas siempre estarán exentas de dificultades como la anterior. Hasta el momento nadie ha sido capaz de entregar una respuesta convincente (Fodor 1998). Peor aún, el futuro de tan relevante pregunta no se vislumbra demasiado alentador.

Ahora bien, la cuestión que nos preocupa principalmente es si nuestra semántica de la función psicopoiética respeta o no el principio de composicionalidad. Para demostrar este hecho consideremos un concepto compuesto general que llamaremos XY, donde X e Y son sus conceptos constituyentes. Analicemos, a continuación, una instanciación cualquiera del concepto XY desde dos perspectivas diferentes. Por un lado, podemos visualizar una instanciación de dicho concepto con éste aportando especificidad respecto de un cierto cúmulo de propiedades (que llamaremos A); es decir, haciendo que el proceso psicopoiético producido por la actitud proposicional en que éste participa dependa necesariamente de dichas propiedades. Por otro lado, podemos visualizar el mismo fenómeno como la instanciación de dos conceptos separados: X e Y. Aplicando la misma definición, tendremos a éstos aportando especificidad respecto de sus respectivos cúmulos de propiedades (que llamaremos B y C, respectivamente). En consecuencia, como el fenómeno es metafísicamente uno, considerado desde dos perspectivas diferentes, ambos aportes de especificidad, el del concepto XY (A) y el de los conceptos X e Y por separado (B y C), deben coincidir ( $A=f(B,C)$ ). La conclusión ineludible es que la función psicopoiética del primero ha de ser una función matemática de las

funciones psicopoiéticas de X e Y; y, por definición, el contenido de XY debe ser también una función matemática de los contenidos de X e Y.

Por otro lado, todo concepto compuesto tiene una estructura sintáctica que en la demostración anterior hemos dejado en la sombra. Ésta, sin embargo, es la que define la “función matemática” en que los contenidos de los conceptos constituyentes se combinan para dar lugar al contenido del concepto compuesto. Con todo, cualquiera sea esta función matemática, esta no cambiará el hecho de que los procesos psicopoiéticos producidos por las actitudes en que el concepto compuesto participa dependerán, de un modo u otro, de las propiedades metafísicas asociadas al contenido de sus constituyentes.

En el ejemplo de serpiente zumbante, tan nocivo para otras semánticas, las propiedades asociadas al contenido de SERPIENTE son, obviamente, las definidas por el objeto genérico “serpiente”; mientras que la asociada al contenido de ZUMBANTE es la propiedad de zumar. Del mismo modo, el contenido del concepto SERPIENTE ZUMBANTE corresponde a su función psicopoiética; esto es, hacer que los procesos psicopoiéticos caracterizados por las actitudes proposicionales en que éste participa dependan necesariamente de un cúmulo de propiedades metafísicas, las cuales –difícil dudarlo- han de identificarse con las propiedades de las serpientes más la propiedad de zumar. Esto se debe a que en la estructura sintáctica del concepto SERPIENTE ZUMBANTE está implícito un conectivo lógico “y” (formalmente denotado por “.”) que define una intersección de los cúmulos de propiedades metafísicas asociados a los conceptos SERPIENTE y ZUMBANTE como el cúmulo respecto del cual el concepto compuesto aporta especificidad.

La razón por la que nuestra semántica no falla donde las semánticas del rol conceptual sí lo hacen es que la función psicopoiética de un concepto no depende de las inferencias en que éste participa. Dos resultados extraíbles de nuestra teoría pueden avalar este hecho. En primer lugar, sólo las actitudes proposicionales iniciales caracterizan a un proceso como siendo específico respecto de un objeto o cúmulo de propiedades. En estricto rigor, no importa el curso particular que siga un proceso psicopoiético una vez puesto en marcha para determinar de qué propiedades dependerá *necesariamente* el éxito o fracaso de dicho proceso; tan sólo importa las actitudes proposicionales que lo produjeron. En segundo lugar, las inferencias que un sujeto puede realizar a partir de ciertas actitudes proposicionales (i.e. de ciertas creencias) son parte del devenir particular del proceso psicopoiético caracterizado por las actitudes proposicionales iniciales que lo produjeron. Consecuentemente, dichas inferencias no

alterarán el objeto o cúmulo de propiedades de que dependerá necesariamente el éxito o fracaso de dicho proceso ni tampoco el contenido de las actitudes que lo caracterizan.

A modo de ejemplo, consideremos el caso hipotético y acaso un tanto ridículo de una persona que está convencida (ya sea porque lo escuchó o por otra razón) de que la carne de las serpientes zumbantes es benigna para la salud. Supongamos que esta persona realizó un viaje para capturar una serpiente de este tipo y que está situada en frente de una en una selva en algún lugar remoto. Las siguientes actitudes proposicionales se instanciarán en la mente de dicha persona: (1) El deseo de capturar una serpiente zumbante; (2) La creencia de que esto (que está frente a ella) es una serpiente zumbante. Ambas caracterizan el proceso psicopoiético que se desarrollará a continuación. Éste puede incluir la inferencia desde “esto es una serpiente zumbante” a “esto es peligroso”, la cual presumiblemente determinará el uso de algún artefacto (e.g. la rama de un árbol) por parte del sujeto para realizar la captura. Sin embargo, el proceso caracterizado por las actitudes proposicionales (1) y (2) que lo produjeron, sea cual sea el curso particular que tome (i.e. haga la captura con una rama o directamente con las manos), tendrá éxito o fracaso dependiendo necesariamente de las propiedades de las serpientes zumbantes. Cuáles crea el sujeto que son las propiedades de las serpientes zumbantes (e.g. ser peligrosa o que su carne es benigna) simplemente no tiene importancia para efectos de la dependencia del proceso. De hecho, puede que el sujeto no hubiese realizado dicha inferencia y el éxito o fracaso del proceso hubiese también dependido necesariamente de las mismas propiedades de las serpientes zumbantes. Lo que el sujeto crea o deje de creer no afecta en nada el contenido de los conceptos pues los procesos a que aquellos dan lugar siguen dependiendo del cúmulo de propiedades que constituye su contenido.

En el fondo de tan complicado argumento se encuentra un hecho que ya fue mencionado anteriormente: las creencias que puedan instanciarse en la mente de un sujeto a raíz de un proceso psicopoiético pueden ayudar a que éste sea exitoso, pero esto no cambia el hecho que, sea o no exitoso, esto dependerá necesariamente de las propiedades que definen al referente. Puesto de otro modo, en nuestra definición de contenido conceptual, no estamos exigiendo a un proceso producido por una actitud proposicional en que participa  $X$  que sea exitoso para ser *acerca de* del objeto  $x$ . Tan solo pedimos que, sea o no exitoso, esto necesariamente se deba a las propiedades metafísicas de  $x$ . En nuestro ejemplo, no importa de qué manera el sujeto intente cazar una serpiente zumbante, si lo hace con sus manos o con una escopeta; no importa, incluso, si está en lo correcto al pensar que la carne de dicha serpiente es benigna

para su salud; lo que importa es que la estabilidad autopoiética de su organismo, una vez instanciado el proceso caracterizado por las actitudes proposicionales (1) y (2), dependerá, para bien o para mal, de las propiedades de las serpientes zumbantes.

Por último, otro hecho muy atinente a la presente temática es la forma como se relacionan los referentes de un conjunto de conceptos para dar lugar al referente de un concepto compuesto por los anteriores. Si bien este principio no ha alcanzado tanta relevancia como el anterior, la mayor parte de las semánticas no-solipsistas concuerdan en que, en general, la referencia de un concepto compuesto debe ser una función (en el sentido matemático del término) de las referencias de sus constituyentes. A esta regla se ha llamado principio de composicionalidad referencial, para distinguirlo del principio de composicionalidad (semántico) recién discutido.

Por medio de un argumento similar al que esgrimimos para demostrar que nuestra semántica de la función psicopoiética respetaba el principio de composicionalidad (semántico), podemos mostrar que también lo hace con el de composicionalidad referencial. En efecto, según ésta, la referencia de un concepto es el(los) objeto(s) que, de existir, sus propiedades metafísicas determinan necesariamente el éxito o fracaso de los procesos psicopoiéticos producidos por las actitudes en que aquél participa. Ahora bien, en el caso de un concepto compuesto, los procesos psicopoiéticos producidos por las actitudes proposicionales en que éste participe dependerán, por definición y dado que los conceptos constituyentes participan también, necesariamente de las propiedades metafísicas de los referentes de éstos últimos. Consecuentemente, el referente de un concepto compuesto debe coincidir con alguna combinación o función matemática de los referentes de sus constituyentes. Análogamente con el caso anterior, ésta estará dada por la sintaxis del concepto compuesto.

En el fondo del asunto, no es extraño que nuestra semántica de la función psicopoiética respete ambos principios de composicionalidad. Más aun, es perfectamente esperable que toda semántica externalista que respeta el principio de composicionalidad semántica, debido a la estrecha vinculación del contenido con la referencia, respete también el principio de composicionalidad referencial. La nuestra no es más que una de éstas donde esta virtuosa vinculación se aprecia nítidamente.

## **Holismo Semántico**

El último obstáculo frente al cual pondremos a prueba nuestra teoría es el llamado holismo semántico. Éste postula que el contenido de un concepto está determinado, de un modo u otro, por las relaciones que éste posee con otros conceptos. Éstas pueden ser de diversos tipos sin perjuicio de que las más comunes candidatas sean las relaciones inferenciales, más o menos asimilables con la red de creencias en que los conceptos participan.

La principal dificultad que el holismo semántico supone es explicar coherentemente la publicidad de los conceptos; esto es, la intuición de que el significado de mi concepto de perro coincide con el de cualquier otro ser humano que lo posea (Fodor 1998). En el caso de aquellas semánticas que postulan las relaciones inferenciales como determinantes del contenido conceptual, esto supone que éste es dependiente de propiedades epistémicas particulares de cada sujeto, violando, así, la publicidad de los conceptos.

Con todo, la popularidad de las semánticas holistas fue hasta hace poco tiempo un hecho indiscutible. La razón probablemente se encuentre en un razonamiento plausible aunque ilegítimo a partir de una premisa plausible y legítima. Si las relaciones inferenciales en que los conceptos participan están determinadas por sus contenidos, más que por la sintaxis de los mismos, entonces presumiblemente el contenido conceptual, a la inversa, estará determinado por las relaciones inferenciales. Por más que el punto de partida sea perfectamente creíble, nada hay de demostrativo en esta deducción. Su plausibilidad exclusivamente analógica termina socavada por las enormes dificultades que el holismo semántico involucra (Fodor y Lepore 1992).

Estos inconvenientes han sido persistentemente acusados por Fodor y otros filósofos. La raíz de los compromisos holistas de las semánticas del rol conceptual se encuentra, en todo caso, en la combinación de dos hipótesis fundamentales en su formulación. Primero, que el contenido conceptual es identificable con la función que el concepto juega en la actividad mental. Segundo, que el modo adecuado de interpretar la noción de función es al estilo Máquina de Turing; esto es, como mapeo de estados mentales en otros estados mentales. Ambas juntas dan por resultado una idea del contenido conceptual como el conjunto de mapeos entre estados mentales –presumiblemente creencias-; en otras palabras, como la estructura de inferencias en que los conceptos participan. Esto, ya está dicho, amenaza la publicidad de los mismos.

Para nuestra teoría la situación es muy distinta pues la noción de función a la que apela es teleológica y no Máquina de Turing. Esto supone concebir la función de un concepto como su aporte a un proceso global parte del cual son las transiciones entre estados mentales. Conforme con esto, entenderemos las instanciaciones de inferencias como un tipo de (si no todas las) transiciones entre estados mentales las que serán, a su vez, consideradas parte del entramado de procesos psicopoiéticos constitutivos de la conducta (ver figura 1, capítulo 4), aun cuando no suficientes para la determinación de la función de un concepto. Por el contrario, la función de un concepto no será concebida en términos de qué estados mapea y qué estados lo mapean sino que ésta estará determinada, según nuestra propuesta, por el contexto completo del sistema autopoiético.

Junto con esto, el que las instanciaciones de inferencias sean consideradas parte de la conducta intencional es un hecho para el cual puede encontrarse abundante evidencia filosófica. Por ejemplo, esto queda de manifiesto en el hecho de que existe un sinnúmero de conductas que prescindan de una expresión motora, limitándose a una secuencia de estados mentales intencionales (Dretske 1991). Aparte del ejemplo del capítulo 4 del presente trabajo podemos citar el sencillo caso de una persona que finge estar muerta. Por cierto que allí habría conducta intencional debido a la mera existencia de transiciones entre estados mentales intencionales, aun cuando ésta carecerá de expresión motora. La conclusión es que de no ser las instanciaciones de inferencias parte de la conducta intencional (viz. parte de los procesos psicopoiéticos) difícilmente éstas podrían constituir conducta intencional por sí solas en algunos casos.

Por otro lado, más allá de las disputas en torno al funcionamiento de la mente, es indudable que los factores que determinan la instanciación de cualquier conducta intencional pueden reducirse a dos: la información almacenada en la mente y la influencia ambiental. La primera de ellas es la relevante en el análisis del problema del holismo. Presumiblemente ésta es identificable con la red de creencias no-instanciadas que posea un sujeto y, en último término, con algún tipo de relaciones interconceptuales. De este modo, a la vez que la red de creencias de los individuos determina la conducta intencional, por definición, determinará también el curso particular que sigan los procesos psicopoiéticos una vez que éstos se han puesto en marcha. Así, por ejemplo, habitualmente las instanciaciones de inferencias que se lleven a cabo al interior de un cierto proceso psicopoiético estarán determinadas por la estructura de creencias subyacente, de modo tal que cada una de las creencias que se instancie en él producirá, a su vez, un nuevo (sub) proceso psicopoiético que se enmarcará en el anterior, en

un esquema similar al de la *figura 1* (capítulo 4). Naturalmente, estas actitudes proposicionales diferirán frecuentemente en contenido, de donde, según nuestra teoría, el éxito o fracaso de los (sub)procesos psicopoiéticos que ellas caracterizan dependerá necesariamente de cúmulos de propiedades metafísicas distintos.

En el caso de Clyde, expuesto en el capítulo 4, el proceso psicopoiético principal que se instanciará estará caracterizado por las actitudes: (1) el deseo de beber cerveza y (2) la creencia de que hay cerveza en el refrigerador. Aun cuando este proceso dependerá *necesariamente* sólo de las propiedades metafísicas de la cerveza, el refrigerador, etc., constitutivas de su contenido; aun así, en la medida que Clyde detecte al perro que está echado en el pasillo que conduce a la cocina, instanciará (3) la creencia de que su perro está en el camino a la cocina y (4) el deseo de eludir su perro. Esto hará que se inicie un nuevo (sub)proceso psicopoiético caracterizado por (3) y (4) diferente en contenido del proceso caracterizado por (1) y (2). El hecho fundamental a entender respecto de esto es que si bien el proceso (3)-(4) dependerá necesariamente de las propiedades metafísicas del perro en cuestión, las cuales constituirán su contenido; el proceso principal (1)-(2) dependerá de éstas sólo contingentemente, de modo tal que esta dependencia no alterará el contenido asociado a las actitudes que lo produjeron.

El hecho de que los (sub)procesos que se vayan desencadenando en el transcurso de un proceso psicopoiético principal no alteren la dependencia modal de éste de las propiedades metafísicas constitutivas del contenido de las actitudes que lo produjeron parece avalar la ausencia de compromisos holistas en nuestra semántica de la función psicopoiética. Y la razón de fondo por la que nos separamos en este aspecto de las semánticas del rol conceptual es que ninguna de las premisas que implican los compromisos holistas de éstas (la hipótesis de que el contenido conceptual se identifica con la función del concepto ni la interpretación matemática de la noción función) son estrictamente compartidas por nuestra teoría. Tal vez, para ser justos, sí compartimos una versión más débil de la primera de ellas. Esta es que el contenido conceptual, identificado aquí con un cúmulo de propiedades metafísicas, está determinado por (si bien no se identifica con) la función del concepto. Claro está, como discutíamos más arriba, bajo una noción de función radicalmente distinta de la denominada Máquina de Turing.

A pesar de esto, aun persiste una intuición respecto de nuestra semántica que hace temer compromisos holistas. Ésta es que la red de creencias determinan de un modo u otro el curso

particular que seguirán los procesos psicopoiéticos una vez en marcha. Esto pareciera indicar que los procesos psicopoiéticos no dependerán sólo de las propiedades metafísicas que constituyen –en la presente teoría- su contenido, lo cual no sólo supondría la amenaza del holismo sino que también podría complicar la determinación del contenido de los conceptos como dependencia de los procesos psicopoiéticos. Sin embargo, en lo que resta veremos que éstos dependen de la red de creencias de una forma distinta que de las propiedades metafísicas constitutivas de su contenido. Consecuentemente, la función psicopoiética no dependerá de aquélla y no estará en peligro la determinación del contenido como tampoco una posible vinculación con el holismo.

En efecto, lo que hace que un concepto particular sea el que es, es su capacidad para hacer que procesos psicopoiéticos distintos, causados por medio de las actitudes en que participa, dependan necesariamente de un mismo objeto (o cúmulo de propiedades). Esta regularidad determina su función y es, pues, lo que permite su individuación funcional. Ésta es también la razón por la que hemos insistido en caracterizar los procesos psicopoiéticos en términos de las actitudes proposicionales que los producen y no de su desarrollo posterior. Ahora bien, no obstante la red de creencias puede determinar el devenir particular de un proceso psicopoiético, esto no cambia el hecho de que todos los procesos psicopoiéticos producidos por actitudes en que participa el concepto PERRO comparten una característica común; a saber, que su éxito o fracaso depende de las propiedades metafísicas de los perros.

Por ejemplo, supongamos que dos personas situadas frente a un mismo perro instancian actitudes proposicionales idénticas, como pueden ser: (1) el deseo de acariciar a un perro y (2) la creencia de que eso (que está frente a ella) es un perro. Supongamos además que, sin embargo, ambas personas difieren en sus creencias respecto de los perros. Una tiene (3) la creencia que los perros pueden volverse repentinamente agresivos, mientras que la otra no (viz. distintas redes de creencias). Probablemente, los procesos psicopoiéticos producidos por las actitudes (1) y (2) sigan cursos distintos en ambas personas, debido a que la primera presumiblemente instanciará (3) inmediatamente después de éstos. En otras palabras, redes de creencias distintas determinan cursos distintos de procesos psicopoiéticos detonados por las mismas actitudes. Sin embargo, cualesquiera que sean los cursos particulares que, en definitiva, sigan los procesos (i.e. cualesquiera redes de creencias), ambos compartirán el hecho de que su éxito o fracaso dependerá de las propiedades de los perros. Por lo tanto su caracterización en términos de las actitudes que los produjeron es admisible como también lo

es la independencia del contenido conceptual de la red de creencias que determina el curso particular de un proceso psicopoiético una vez instanciado.

Por otro lado, las creencias particulares que se instancien una vez que un proceso psicopoiético se ha puesto en marcha determinarán su curso. Más aun, dichas creencias darán lugar a otros (sub)procesos psicopoiéticos que se enmarcarán en el primero en un esquema como el que figura en el capítulo 4 (figura 1). Sin embargo, no sería apropiado hablar de dependencia de dicho proceso de las creencias que en él se instancien pues éstas serán parte del mismo y, de perogrullo, todo proceso –psicopoiético o no- depende de sí mismo. Más bien, lo realmente relevante de un proceso tal es que, una vez en marcha, su éxito o fracaso dependerá necesariamente de cierto cúmulo de propiedades metafísicas (viz. su contenido) externas al proceso mismo, sólo en virtud de las condiciones que lo produjeron (i.e. las actitudes proposicionales que los caracterizan) sin importar, reitero, cuál sea su curso particular. Esto es lo maravilloso de los conceptos, que hacen que una gran diversidad de procesos psicopoiéticos sean específicos respecto de un mismo objeto (o cúmulo de propiedades).

Con todo, para respetar la idea de que la mente/cerebro es un órgano surgido de la evolución y que el aparato cognitivo posee un alto poder adaptativo, es plausible pensar que la red de creencias que poseen los individuos determina cursos de procesos psicopoiéticos con mayor probabilidad de éxito. Esto podría ser interpretado como una prueba de la dependencia del éxito o fracaso de los procesos psicopoiéticos de la red de creencias subyacente. Sin embargo, nada más lejos de la realidad pues una cosa es causar el curso particular de un proceso y otra muy distinta es la dependencia de éste en aquello que lo causa. La red de creencias causa el curso que seguirá un proceso psicopoiético pero no es causa de la puesta en marcha de los procesos psicopoiéticos. Las actitudes propocisionales, por su parte, causan que los procesos psicopoiéticos se pongan en marcha. Las propiedades metafísicas del referente de las actitudes, por último, causan que un proceso psicopoiético, sea cual sea el curso que haya seguido, tenga éxito o fracase. Esto es lo que queremos indicar cuando hablamos de dependencia de los procesos de las propiedades metafísicas del referente. En consecuencia, existe una relación entre las actitudes que desencadenan la puesta en marcha de los procesos psicopoiéticos y las propiedades de las que dependerá el éxito o fracaso de los mismos la cual es independiente del curso particular del proceso. Esta es, según proponemos, la relación de intencionalidad.

Por otro lado, convengamos que esta relación entre las actitudes proposicionales y las propiedades metafísicas con independencia del curso particular del proceso detonado por dichas actitudes no resulta rara cuando la traducimos al lenguaje de la Psicología Intencional. De hecho, lo anterior corresponde a una aseveración completamente trivial en dicho lenguaje; a saber, que un mismo par de actitudes proposicionales puede dar lugar a diversas conductas sin que se altere el contenido de las primeras. Nada de extraño hay en esto, como tampoco lo hay en que distintos procesos psicopoiéticos producidos por las mismas actitudes dependen de las mismas propiedades metafísicas.

En resumen, el contenido de un concepto no depende, en nuestra teoría, de la red de creencias en que éste participa. Así, no tenemos compromisos holistas y la publicidad del contenido conceptual queda garantizada. Es factible pensar, en todo caso, que la estructura de creencias particular de cada individuo ayude a orientar desarrollos de los procesos psicopoiéticos con mayor probabilidad de éxito, pero, repito, esto nada importa al hecho de que todos cuantos poseen el concepto PERRO instanciarán actitudes proposicionales que darán lugar a procesos psicopoiéticos que dependerán de las propiedades metafísicas de los perros independientemente del curso particular que sigan. Éste estará determinado por la red de creencias. La relación entre las propiedades metafísicas y las actitudes proposicionales, en cambio, constituirán, precisamente, lo que llamamos intencionalidad, la cual será independiente del curso particular del proceso y, en definitiva, de la red de creencias que el sujeto alberga.

Finalmente, resulta interesante establecer que lo anterior supone también un importante corolario respecto de las condiciones de posesión de conceptos. Éste es que las condiciones de posesión de un concepto no incluyen el conocimiento (acceso epistémico) por parte de un individuo del cúmulo de propiedades metafísicas que, según nuestra teoría, constituyen su contenido. A primera vista, éste podría ser el corolario más extraño que se sigue de la presente teoría. Sin embargo, en una mirada un poco más minuciosa, el asunto resulta acaso natural. De hecho, sólo corresponde a decir que el contenido conceptual no es algo aprehensible epistémicamente sino metafísicamente. En otras palabras, el contenido conceptual no es una categoría epistémica sino metafísica. Esto es, justamente, lo que distingue nuestra propuesta de las semánticas del rol conceptual.

## Bibliografía

- Beckner, M. 1959. *The Biological Way of Thought*. New York.
- Block N. 1978 *Troubles with Functionalism*; Perception and Cognition, vol.IX. University of Minnesota Press.
- Borst, C.V.(Ed.) 1983 *The Mind/Brain Identity Theory*; Macmillan Press.
- Brentano, F. 1874 *Psychology from an empirical point of view*, Versión Inglesa por A.C. Rancurello, D.B. Terrell, and L. McAlister Brentano, F., London: Routledge, 1973.
- Burge, T. *Individualism and the Mental*; En French P., Euhling, T., y Wettstein, H. (Eds.), *Studies in Epistemology*. Vol 4., University of Minnesota Press.
- Canfield, J 1964. *Teleological Explanation in Biology*. *British Journal for the Philosophy of Science*, 14.
- Chuaqui, B. 1995 *Sobre el Reduccionismo en las Ciencias*, *Revista Médica de Chile*, Vol.123, n°10 (1995).
- Churchland, P.M. 1981 *Eliminative Materialism and Propositional Attitudes*; *The Journal of Philosophy* 78.
- Churchland P.S. & Sejnowski T.J. 1990 *Neurophilosophy and Connectionism*; en Lycan (Ed.) 1990.
- Dennett, D. 1975. *Why the Law of Effect Will Not Go Away*. *Journal of the Theory of Social Behavior*.
- Dennett, D. 1981 *True believers: the intentional strategy and why it works*; Oxford University press. En lycan 1990.
- Devitt, M. 1981 *Designation*; Columbia University press
- Dretske, F. 1981 *Knowledge and the Flow of Information*; MIT Press.
- Dretske, F. 1991 *Explaining Behavior*; MIT Press.
- Fodor, J. 1975 *The Language of Thought*, Thomas Y. Crowell Co.
- Fodor, J. 1987 *Psychosemantics*; MIT Press.
- Fodor, J. & Lepore, E. 1992 *Holism: A Guide's Shopper*. Oxford: Blackwell.
- Fodor, J. 1998 *Concepts: Where Cognitive Science went wrong*; MIT Press.
- Frege G. 1892 *On sense and reference*; Versión Inglesa en P. Geach and M. Black (Eds. and trans.), Oxford: Blackwell, third edition, 1980.
- Gallistel, C.R. 1980. *The Organization of Action: A New Synthesis*; Hillsdale, N.J. : Erlbaum.
- Harman G. 1987 *(Nonsolipsistic) Conceptual Role Semantics*, "New Directions in Semantics", edited by Ernest LePore, London, Academic Press.
- Laudan L. 1981 *A Problem-Solving Approach to Scientific Progress* en Hacking, I. (Ed.), *Scientific Revolutions*, (Oxford Readings in Philosophy Series, O.U.P., Oxford), 144-55.

- Loar, B. 1983 *Must beliefs be sentences?* En P.D. Asquith & T. Nicles (Eds.) Vol.2 . East Lansing MI: philosophy of science association.
- Lycan, W.G. 1987. *The Continuity of Levels of Nature*. En Lycan (Ed.) 1990.
- Lycan W. (Ed.) 1990 *Mind and Cognition*; Basil Blackwell.
- Maynard Smith, J. and Szathmáry, E. 2001 *Ocho Hitos de la Evolución*; Tusquets.
- Maturana, H. & Varela, F. 1964. *De Máquinas y Seres Vivos*; Universitaria.
- Nagel, E. 1961. *The Structure of Science*; New York.
- Putnam H. 1975 *The meaning of 'meaning'* ; en Gunderson, K. (Ed.) University of Minnesota Press.
- Quine, W.O. 1953 *Two Dogmas of Empirism*; en *From a Logical Point of View* Cambridge University Press.
- Russell, B. 1974 *La Filosofía del Atomismo Lógico*. Londres (s.n.)
- Sherrington, C. 1947 *The Integrative Action of the Nervous System*. Cambridge 1947.
- Sober, E. 1985. *Putting the Function Back into Functionalism*. En Lycan (Ed.) 1990.
- Stampe, D. 1977 *Toward a Causal Theory of Linguistic Representation*; Midwest Studies in Philosophy, II.
- Stich, S. 1983 *From Folk Psychology to Cognitive Science*; MIT Press.
- Van gulick R. 1980 *Functionalism, information and content*; en *Nature and System 2*.
- Von Eckardt, B. 1995 *What is Cognitive Science?*; MIT Press.
- Wittgenstein, L. 1921 *Tractatus Lógico-philosophicus*; Versión inglesa 1922, Routledge & Kegan Paul.
- Wright, L. 1973. *Functions*. *Philosophical Review*, 82.