

## REALISMO INTENCIONAL, ELIMINATIVISMO Y PSICOLOGÍA CIENTÍFICA<sup>1</sup>

LIZA SKIDELSKY  
Departamento de Filosofía  
Universidad de Buenos Aires  
lskidels@filo.uba.ar

### Resumen

El realismo intencional acérrimo (RIA o teoría representacional/computacional de la mente de J. Fodor) quiere dos cosas que, según intentaré mostrar, son difíciles de lograr conjuntamente: estados (1) con contenido semánticamente interpretable y (2) que tengan rol causal en la producción de la conducta (y de otros estados intencionales). Si bien no hay dificultades para obtener(1), el problema es con (2). La estrategia del RIA consiste en postular la existencia de leyes intencionales causales. El problema es que esas leyes son de un estatus dudoso, al punto que muchos consideran que no son posibles y ni siquiera necesarias. Según el RIA, si no se puede obtener conjuntamente (1) y (2) se desprenderían dos consecuencias indeseables: si los estados intencionales no tienen rol causal *qua* intencionales significaría que no se es realista acerca de las creencias y deseos; y que no sean posibles las leyes intencionales implicaría que no se puede tener una psicología científica. En la primera parte del trabajo me ocuparé de exponer las tesis del RIA y mostraré que no puede cumplir conjuntamente con las exigencias (1) y (2). En la segunda parte me ocuparé de defender la idea de que del hecho de que no se pueda conciliar (1) y (2) no se sigue un eliminativismo acerca de los estados intencionales ni la imposibilidad de que la psicología sea una ciencia.

PALABRAS CLAVE: modelos computacionales, actitudes proposicionales, rol causal, leyes intencionales.

### Abstract

The Hairy Intentional Realism (HIR or J. Fodor's Computational/Representational Theory of Mind) aims for two things that I shall try to show are difficult to jointly obtain: states (1) with an interpretable semantic content, and (2) with causal role in

<sup>1</sup> Una versión anterior de este trabajo con el título "Estados intencionales y eliminativismo" fue leída en el VI Coloquio Internacional Bariloche de Filosofía, Argentina, 2002, y en el seminario del proyecto UBACyT F003: "Conceptos, categorías y significado", 2002. Agradezco a los participantes de ambas reuniones, en especial a Eduardo Rabossi, Ma. Cristina González, Estela Santilli y Ana Hulton, cuyos comentarios me permitieron mejorar el trabajo, como también a Diana Pérez y Karina Pedace quienes leyeron versiones anteriores.

the production of behavior. There is no difficulty in obtaining (1), the problem is with (2). HIR's strategy consists in the postulation of causal intentional laws. The problem is that these laws have a doubtful status to the point that many consider them not possible and not even necessary. According to HIR, if (1) and (2) cannot be jointly obtained, two unwanted consequences emerge: if intentional states do not have a causal role *qua* intentional, that would mean ceasing to be realistic about them; and if intentional laws are not possible that would imply that we could not have a scientific psychology. In the first part of this paper I shall exhibit HIR's main theses and I shall show that it cannot accomplish the two demands. Second, I shall defend the idea that because of the fact that (1) and (2) cannot be reconciled, neither an eliminativism about intentional states, nor the impossibility that psychology is a science can follow.

KEY WORDS: computational models, propositional attitudes, causal role, intentional laws.

El realismo intencional quiere dos cosas que, según intentaré mostrar, son difíciles de conciliar: estados (1) con contenido semánticamente interpretable y (2) que tengan rol causal en la producción de la conducta (y de otros estados intencionales). La primera exigencia viene dada por la postulación de estados intencionales como aquellos estados mentales cuyo contenido se expresa en términos semánticos/intencionales. El problema surge cuando se exige que esos estados tengan un papel causal en la producción de la conducta *en virtud de* su contenido semántico. La estrategia del realismo intencional acérrimo [RIA], del cual me ocuparé en este trabajo, consiste en postular la existencia de leyes intencionales causal es. Así, mientras haya leyes causales que subsuman propiedades semánticas/intencionales, los estados intencionales tendrán poder causal. El problema es que esas leyes son de un estatus dudoso, al punto que muchos consideran que no son posibles y ni siquiera necesarias.

Ahora bien, si se acepta que no hay leyes intencional es, entonces, según el RIA, el contenido de los estados intencionales no tiene rol causal. Esto es, no se puede cumplir con el requisito (2). Según el RIA, si no se puede conciliar (1) y (2) se desprenderían dos consecuencias indeseables: (a) si los estados intencionales no tienen rol causal *qua* intencionales significaría que no se es realista acerca de las creencias y deseos (dado que sólo lo que tiene poder causal es real), y (b) que no sean posibles las leyes intencionales implicaría que no se puede tener una psicología científica (dado que toda ciencia requiere de leyes). En la primera parte del trabajo me ocuparé de exponer las tesis del RIA y mostraré que no puede cumplir conjuntamente con las exigencias (1) y (2). En la segunda parte me ocuparé de defender la idea de que del hecho de que no se pueda

conciliar (1) y (2) no se sigue un eliminativismo acerca de los estados intencionales ni la imposibilidad de que la psicología sea una ciencia.

## 1. El realismo intencional acérrimo quiere demasiado

### 1.1 Realismo intencional acérrimo

Todo realista intencional considera que hay estados mentales que son “acerca de algo”. Por ejemplo, la creencia de Juan de que el pasto es verde es acerca del pasto y el deseo de Juan de tomar una cerveza es acerca de la cerveza. Los contenidos de esos estados son (1) semánticamente interpretables y (2) poseen un papel causal en la producción de la conducta. Que sean semánticamente interpretables significa que tienen propiedades semánticas, por ejemplo, condiciones de satisfacción: el contenido de las creencias tiene condiciones de verdad y el de los deseos tiene condiciones de cumplimiento. El contenido de la creencia de Juan de que el pasto es verde será verdadero o falso según que el pasto sea verde o no. Y el contenido del deseo de Juan de tomar una cerveza se cumplirá o no según si Juan toma la cerveza o no. Las propiedades semánticas permiten hacer generalizaciones puesto que éstas últimas se definen sobre elementos que tienen valores de verdad. Así, las generalizaciones que involucran contenidos figuran en el *explanans* de las explicaciones de la conducta de los individuos. De esta manera, se puede explicar (o predecir) la conducta de los individuos (incluido Juan) de comprar sillas verdes para su jardín porque “creen que el pasto es verde” y “desean que sus sillas tengan el mismo color que el pasto”. Esto se conecta con la idea de que los estados intencionales poseen un papel causal en la producción de la conducta en el sentido de que estos estados son condición suficiente para producir un cierto comportamiento. Así como explicamos la rotura de un frasco de vidrio con agua cuando se lo calienta apelando al factor causal del momento del impacto de la colisión de las moléculas del agua con la superficie del frasco, podemos explicar la conducta de Juan al tomar una cerveza en el bar apelando al factor causal de la conjunción de su deseo de tomar una cerveza y su creencia de que hay cerveza en el bar.

Cualquier filósofo que postule la existencia de estos estados se compromete con estas dos tesis. Incluso, se dice que no necesitamos ser filósofos para suscribirlas porque están implícitas en nuestros intercambios cotidianos con las personas. Así, se dice que estos in-

tercambios suponen una psicología de sentido común (PSC, lo que se suele llamar en inglés *folk psychology*) que nos sirve para explicar y predecir nuestro propio comportamiento y el de los demás. Todo realista intencional acepta esta caracterización de nuestros estados intencionales de la PSC. A partir de aquí, las posturas realistas ofrecen distintos grados de realismo, según se especifiquen (1) y (2). Desde el “realismo moderado” de Dennett hasta el “realismo fuerte industrial” de Fodor (en mis términos, RIA), pasando por el “realismo fuerte normal” de Davidson<sup>2</sup>. En lo que sigue me ocuparé de exponer algunas de las tesis principales del RIA formuladas en la teoría representacional/computacional de la mente [TRCM] de Jerry Fodor, que intenta incorporar la PSC a una psicología científica, porque considero que es en este tipo de realismo donde es difícil congeniar justamente lo que un defensor del RIA quisiera tener, esto es, (1) y (2)<sup>3</sup>.

Se suele llamar a los estados intencional es actitudes proposicionales [AP] porque expresan una relación (de creencia, deseo, etc., de ahí “actitud”) entre un individuo y una proposición (que es el contenido de la creencia). Según la TRCM y siguiendo nuestro ejemplo, para el evento que consiste en la creencia de Juan de que el pasto es verde hay un evento correspondiente que consiste en que Juan está relacionado con (tiene en su “caja de creencias”) una instancia de una representación mental [RM] que tiene el contenido “el pasto es verde”. Si estar en un estado intencional significa que hay un evento en el que estamos relacionados con RMs, lo que nos interesa es la naturaleza de esas RMs.

Las RMs son símbolos que poseen propiedades semánticas y sintácticas. Entre las primeras está la propiedad de “tener un contenido” y entre las segundas, la propiedad de “ser expresadas por medio de modos de presentación”. Así, una RM desde un punto de vista semántico tiene un significado expresado en su contenido y desde el punto de vista sintáctico está expresada de cierta forma. Por ejemplo, las RMs de AGUA y H<sub>2</sub>O tienen el mismo contenido

<sup>2</sup> La clasificación es de Dennett (1991). Extraigo la expresión “acérrimo” de la frase de Fodor (1999, p. 42, n. 3): “soy un Realista Acérrimo”.

<sup>3</sup> Quizá haya otras formas de realismo (en especial, materialismos no reduccionistas) que son susceptibles de la misma dificultad. En particular se ha acusado a Davidson de no lograr (2) en razón de cómo despliega (1) (esto es, en especial, el monismo anómalo). Pero aun cuando sea el caso, de todas formas los argumentos que utilizaré apuntan en particular a una teoría de la mente que es representacional y computacional, características que no posee la teoría de Davidson.

porque significan lo mismo (tienen el mismo referente), pero distintos modos de presentación porque la formas de las expresiones son distintas (tienen distintas combinaciones de formas de letras). En otras palabras, los modos de presentación son los vehículos sintácticos de los contenidos semánticos. Así, Juan puede tener una inscripción de la RM AGUA mientras que Pablo tiene una inscripción de la RM H<sub>2</sub>O, pero comparten el mismo contenido (la misma proposición). De manera que las APs son, en realidad, relaciones de tres lugares entre un organismo, una proposición y un modo de presentación<sup>4</sup>.

Una vez desplegada la noción de AP, nos ocuparemos de los procesos mentales. Estos consisten en secuencias causales/computaciones sobre RMs. Una cadena de pensamientos es una secuencia causal de instancias de RMs. Pero también los estados intencionales están relacionados causalmente con inputs sensoriales y outputs motores. Así, los contenidos producen estados conductuales caracterizados intencionalmente. La única manera de que los contenidos de las RMs sean cognitivamente relevantes para explicar la conducta es que formen parte de la estructura causal del sistema. Para ello, las representaciones tienen que estar físicamente instanciadas de manera de poder afectar a un sistema físico. Ahora bien, las computaciones sólo son sensibles a la sintaxis (o forma, o estructura) de las RMs, no a su contenido semántico (lo que significan los símbolos)<sup>5</sup>. Lo que aparentemente resulta una dificultad, en realidad es la solución que Fodor encuentra para conectar la semántica de las RMs con la causalidad. Dado que las relaciones causales respetan, de manera fiable, las relaciones entre los contenidos, la sintaxis permite conectar la causalidad con la semántica. Así, la causalidad es sensible a la sintaxis, la sintaxis es sensible a la semántica y por ende, la causalidad es sensible a la semántica.

Cuando Fodor se refiere a la relación entre la sintaxis y la semántica de las RMs, utiliza expresiones del tipo: la sintaxis “respetar” la semántica (Fodor, 1987, pp. 32-33), hay un “paralelismo” entre sintaxis y semántica (Fodor, 1987, p. 34), las relaciones semánticas entre los símbolos pueden ser “imitadas” por sus relaciones sintácticas (Fodor, 1987, p. 41). La idea es la siguiente:

<sup>4</sup> Las RMs conforman un lenguaje del pensamiento, Para una exposición completa del lenguaje del pensamiento y, en general de la TRCM, véase Fodor (1987), cap. 1; (1994), cap. 1, y (1999), cap. 1.

<sup>5</sup> Cf, la “condición de formalidad” de Fodor (1980).

Con todo, la máquina está diseñada de tal manera que transformará un símbolo en otro si y sólo si las proposiciones expresadas por los símbolos que se transforman de esa manera guardan ciertas relaciones semánticas; por ejemplo, la relación que mantienen las premisas con la conclusión en un argumento válido. Semejantes máquinas –los ordenadores, por supuesto– son precisamente entornos en los que la sintaxis de un símbolo determina su papel causal de una manera que respeta su contenido. Esta es, según pienso, una idea estupenda, sobre todo porque funciona (Fodor, 1987, p. 41).

Es así como la TRCM conjuga la PSC con la idea de Turing de que pensar es computar, para lograr una psicología científica. Para que la conducta sea explicada (o predecida) en función de estados intencionales, las APs de la PSC se despliegan como relaciones con RMs instanciadas físicamente (objetos físicos con propiedades semánticas) a las que se aplican computaciones al estilo Turing. Esto permite mantener conjuntamente que los estados intencional es (1) poseen un contenido interpretable semánticamente y que (2) ese contenido tiene un papel causal en la producción de la conducta y de otros estados mentales.

## 1.2 Los argumentos eliminativistas

La estrategia en este apartado consistirá en presentar dos argumentos eliminativistas en contra de las APs. El primero apela a modelos computacionales conexionistas y la conclusión será que las APs, tal como están especificadas por la TRCM, no tienen un rol causal en los modelos conexionistas. El segundo apela a una cuestión metodológica de la TRCM y la conclusión será que en la TRCM las APs no tienen rol causal. De estas conclusiones se suele pretender extraer una adicional que consiste en el eliminativismo fuerte de las APs en los modelos computacionales en general, o al menos, el eliminativismo débil. Mi postura será que estos argumentos no apoyan la conclusión adicional, esto es, no apoyan ninguno de los dos tipos de eliminativismo (apartado 2.1), pero sí son argumentos de peso para extraer la conclusión general de que en los modelos computacionales las APs no tienen un rol causal (apartado 1.2)<sup>6</sup>.

<sup>6</sup> Que se siga esta conclusión general depende de un examen exhaustivo de todas las propuestas computacionales para la arquitectura mental. Como no es posible agotar los modelos existentes, me contento con la conclusión más débil que abar-

### 1.2.1 Argumento que apela a modelos conexionistas

Los argumentos eliminativistas en general suelen estar dirigidos a las APs de la PSC. Se suele aceptar que la PSC es una teoría y estos argumentos pretenden mostrar que es una teoría falsa. Como creo que la PSC no es una teoría, voy a redireccionar el argumento eliminativista que hace uso de modelos conexionistas mediante el siguiente argumento: (i) puesto que las características que los argumentos eliminativistas (basados en modelos conexionistas) les atribuyen a las APs de la PSC son las mismas que Fodor le atribuye a las APs en la TRCM, y dado que la PSC está incorporada a la TRCM (tal como Fodor quiere<sup>7</sup> y como se explicitó anteriormente), se sigue que las APs de la PSC son el mismo tipo de entidad que las de la TRCM, y (ii) puesto que la tesis básica de la TRCM, esto es computación sobre símbolos, es la misma que la del “cognitivismo clásico”, ya que como hemos visto Fodor se inspira en el computacionalismo clásico de Turing, (iii) es lícito reinterpretar los argumentos eliminativistas como una confrontación entre dos tendencias en inteligencia artificial: el conexionismo y el cognitivismo clásico (o TRCM). Así, lo que pretenden establecer los argumentos eliminativistas es que si los modelos conexionistas tienen éxito y en ellos las APs no tienen rol causal, entonces las APs que postula el cognitivismo clásico o bien no existen, o bien no formarán parte de una psicología científica<sup>8</sup>.

Es ampliamente aceptado que hay dos tipos de tesis eliminativistas. El *eliminativismo débil* se formula de dos maneras, una sostiene que las APs no formarán parte de la ontología de la psicología científica (Stich, 1992), la otra sostiene que los estados caracterizados intencionalmente no jugarán un rol explicativo en la psicología científica (Egan, 1995). El *eliminativismo fuerte* sostiene que no existen las AP. Ramsey, Stich y Garon en un artículo de 1990 sostienen un eliminativismo fuerte basado en el argumento de que si cierta clase de hipótesis conexionistas resultan verdaderas, entonces también es verdadero el eliminativismo de las APs. La estrategia consiste en mostrar que las APs poseen ciertas propiedades y dado que en los mo-

---

ca sólo a las dos tendencias más fuertes y desarrolladas que compiten por la explicación de la arquitectura cognitiva: cognitivismo clásico y conexionismo.

<sup>7</sup> Cf. Fodor (1987), p. 37.

<sup>8</sup> Voy a seguir usando la expresión PSC simplemente para mantener fidelidad a la formulación original de los argumentos basados en el conexionismo, pero cada vez que aparezca PSC a lo largo de este trabajo, léase cognitivismo clásico o TRCM.

delos conexionistas no aparecen entidades con esas propiedades, entonces las APs no existen. Ramsey, Stich y Garon (1990) sostienen que las APs se caracterizan por poseer “modularidad proposicional” que consiste en tres propiedades: (a) son funcionalmente discretas: se puede hablar de agentes que obtienen o pierden creencias individuales, (b) son interpretables semánticamente: esto permite realizar generalizaciones nomológicas en virtud del contenido semántico de las APs. En este sentido, los predicados (tal como “cree que p”) involucrados en esas generalizaciones son proyectables porque figuran en generalizaciones y permiten agrupar clases naturales (por ejemplo, todos los que creen que p), y (e) son causalmente eficaces: producen otras APs y conducta. Dado que en los modelos conexionistas no hay ninguna entidad que posea estas propiedades, Ramsey, Stich y Garon (1990) concluyen que las APs no existen.

Este argumento ha tenido una serie de réplicas, la que resulta pertinente es la que proviene de un teórico conexionista. Smolensky (1995) no acepta el consecuente del condicional de Ramsey, Stich y Garon (1990) porque sostiene que si ciertas hipótesis conexionistas resultan verdaderas, entonces la PSC no está “seriamente” equivocada, sino tan sólo “un poco” equivocada, de manera que las APs pueden formar parte de una teoría psicológica científica<sup>9</sup>. Si se analiza el comportamiento de una red conexionista en un nivel superior, es decir en términos de patrones de activación (y no en el nivel inferior de las unidades y pesos de las conexiones individuales), es posible adscribirle creencias a la red entendidas como disposiciones a producir patrones de activación. Los patrones de activación son los que codifican las creencias y éstos tienen interpretación semántica (propiedad (b) porque es en el nivel de los patrones de activación donde se definen las propiedades semánticas de un sistema (en este sentido se podrían proyectar predicados como “creer que p” en generalizaciones que abarquen sistemas cognitivos que creen que p). Además, si bien las creencias no son físicamente discretas porque se especifican a partir de una subregión (del espacio de vectores) de la red, pueden ser funcionalmente discretas (propiedad (a)) porque hay una manera abstracta de identificar esa subregión.

<sup>9</sup> Considero que esta interpretación hace más justicia a la postura de Smolensky que la interpretación de Stich y Warfield (1995) quienes consideran que si se despliega el condicional en dos partes: (i) si ciertas hipótesis conexionistas resultan verdaderas, entonces la PSC está seriamente equivocada, y (ii) si la PSC está seriamente equivocada, entonces se sigue el eliminativismo de las APs, Smolensky acepta (i) y niega (ii). En mi opinión, no acepta ni (i) ni (ii).

Sin embargo, según Smolensky (1995), lo que tiene rol causal en las redes conexionistas no son los patrones de activación sino las unidades y los pesos individuales del nivel inferior que son sobre los que operan los algoritmos. Que algo tenga rol causal es una propiedad de ciertas clases de algoritmos (*nivel algorítmico*) y el nivel de interpretación de qué hace un sistema (*nivel computacional*) no tiene que ser isomórfico con el nivel de cómo lo hace. Así, el conexionismo conserva dos propiedades clásicas de las creencias en el nivel computacional (o semántico o funcional) de descripción de la función que realiza un sistema, pero abandona la causalidad clásica (propiedad (c)) en el nivel algorítmico, es decir, las creencias no tienen un rol causal en la conducta de las redes.

Puesto que las propiedades de las APs que interesan en este trabajo son las propiedades (b) y (c), o lo que es lo mismo (1) y (2), voy a dejar de lado la propiedad (a) y voy a asumir que la respuesta de Smolensky es correcta y que los estados conexionistas tienen (1) una descripción en términos semánticos, pero no tienen (2) rol causal *qua* semánticos, y veamos qué ocurre en el caso de la arquitectura clásica defendida por la TRCM.

### 1.2.2 Argumento que apela a la metodología del cognitivismo clásico

Hemos visto que las APs en los modelos conexionistas no tienen rol causal, a continuación veremos que las APs tampoco tienen rol causal en el cognitivismo clásico (o TRCM). En los modelos computacionales clásicos es posible obtener una descripción de los estados de un sistema computacional en términos semánticos, pero esos estados caracterizados así no cumplen un rol causal *qua* semánticos. Veamos en qué consiste un sistema computacional cognitivo.

Un sistema cognitivo instancia computacionalmente funciones cognitivas, así las capacidades cognitivas consisten en instanciar una función que relaciona contenidos proposicionales porque toma contenidos como argumentos y valores, y los relaciona de una manera que respeta restricciones epistémicas (rationales), a la manera en que se relacionan las premisas con la conclusión. El sistema es cognitivo sólo si puede haber una descripción semántica de la función que realiza. Por ejemplo, que un sistema computacional sea una máquina de sumar radica en que puede ser descripto por una

“función +”:  $+( \langle n, m \rangle ) = s$ <sup>10</sup>. Esta “función +” toma como argumentos (o inputs) la secuencia de botones apretados:  $\langle C, N, +, M, = \rangle$ , esto es,  $\langle \text{despejar, primer adendo, más, segundo adendo, igual} \rangle$  y como valores (o outputs) los numerales exhibidos en una pantalla  $P$ :  $n + m$ . La suma relaciona números, eso significa que el sistema físico no satisface literalmente la “función +” sino que lo que hace es instanciar la “función +” satisfaciendo la “función  $g$ ” (los argumentos y valores de  $g$  son estados físicos del sistema). Bajo interpretación  $I$ , la función  $g$  es una instanciación de la función +.

Las representaciones se definen a partir de la relación inducida por la proyección interpretativa entre los elementos de la función física y la función cognitiva instanciada. Así, la noción de representación tiene sentido porque hay simulación (la instanciación es un caso especial porque relaciona objetos abstractos y sus instancias físicas) e interpretación (la proyección de procesos computacionales sobre otros procesos). Cummins (1989, p. 129) define representación de la siguiente manera: “[una representación]  $R$  representa  $x$  en [el sistema computacional]  $S$  si hay funciones  $g, f$ , y una interpretación  $I$  tal que  $S$  satisface  $g$ ,  $g$  simula  $f$  bajo  $I$  e,  $I(R) = x$ .” Así, los objetos de la computación vistos bajo interpretación son representaciones. Tal como dice Cummins:

Es el hecho de que una función simula a la otra bajo una interpretación fija lo que hace posible pensar en los argumentos y valores de una función como representando los argumentos y valores de la otra. La estructura causal de una máquina de sumar –el hecho de que ejecuta un programa adecuado y por eso satisface la función  $g$ – garantiza que los argumentos y valores de  $g$  le siguen el rastro a los números (1989, pp. 95-96).

El ejemplo de la máquina de sumar es muy simple puesto que el dominio de lo representado no corresponde a un entorno natural. Es distinto el caso de sistemas (como nosotros) con capacidades cognitivas complejas que están insertos en entornos físicos/sociales. En este caso, el contenido semántico de los estados del sistema está determinado por su entorno<sup>11</sup>. Mientras que los estados computacio-

<sup>10</sup> El ejemplo es de Cummins (1989), p. 89.

<sup>11</sup> La tesis externalista en relación con los contenidos establece que el contenido de los estados intencionales de una persona depende de sus relaciones con el mundo. Hay distintas maneras de formular esta tesis, para una formulación en términos de superveniencia, véase Kim (1994).

nales se individúan en términos formales por la función matemática que computan, los contenidos se individúan por sus relaciones con el entorno<sup>12</sup>. Tal es así que no toda teoría computacional es una teoría cognitiva (en el sentido de tener una descripción en términos semánticos/intencionales). Las descripciones matemáticas de un sistema son independientes del entorno externo (dos mecanismos que computan la misma función matemática son el mismo mecanismo aunque estén en dos entornos distintos) y del entorno interno (dos mecanismos idénticos podrían intercambiarse en un sistema sin alteración del mismo). En el caso de los estados intencionales la caracterización de un sistema computacional como cognitivo/intencional depende de que esté inserto en un entorno adecuado que haga posible que sus estados internos se describan como poseyendo cierto contenido representacional.

Según lo que hemos visto, los estados computacionales no tienen sus propiedades semánticas de manera esencial o intrínseca sino que hay una función de interpretación extrínseca que proyecta en estructuras del sistema elementos de un dominio representado. Así, lo que permite la función de interpretación al asignar contenido a los estados de un sistema computacional es poder rastrear los intercambios causal es, pero esos contenidos no tienen ningún rol causal en la secuencia que va de un estado a otro del sistema porque como hemos visto (en el apartado 1.1) los procesos computacionales son sensibles a propiedades sintácticas de los símbolos que manipulan y no a sus propiedades semánticas<sup>13</sup>.

En este apartado 1.2, hemos visto dos argumentos bien fundados por los cuales las APs de los modelos computacionales de la

<sup>12</sup> Se podría objetar que las descripciones matemáticas son semánticas, pero lo que está en juego aquí es la semántica entendida como adscripción de contenidos intencionales. A lo largo de este trabajo no he hecho una distinción entre los términos “semántico” e “intencional” (tal como es la costumbre generalizada en filosofía de la mente).

<sup>13</sup> He seguido algunos de los desarrollos de Cummins (1989) y Egan (1995) para introducir los modelos computacionales clásicos, pero cabe advertir que la conclusión acerca de la impotencia causal de la semántica no se encuentra en sus trabajos, Egan (1995) no se ocupa del rol causal de los estados intencionales (parece darlo por sentado) y Cummins (1989, cap, 10) considera que si hay generalizaciones causales que apelan al contenido semántico, y éstas no son reducibles a generalizaciones que no apelan al contenido, entonces el contenido tiene eficacia causal. Claro que estas generalizaciones causales, tal como él mismo advierte, no son “literales” en el sentido de ser generalizaciones de afirmaciones causales singulares. Conceder esto es todo lo que se requiere para aceptar la conclusión fuerte acerca de la ineficacia causal del contenido *per se*, que se ha extraído en este apartado.

mente (al menos los dos modelos de la inteligencia artificial que hoy compiten por la explicación de la arquitectura mental) si bien cumplen (1), no pueden cumplir con (2).

### 1.3 La solución del realismo intencional acérrimo

A continuación veremos la respuesta de Fodor a la objeción que en la TRCM los contenidos semánticos no tienen rol causal, y las críticas a esta respuesta.

#### 1.3.1 La propuesta de leyes intencionales causales *ceteris paribus*

Habíamos dicho que según la TRCM la semántica está conectada con la causalidad a través de la sintaxis, y eso le permite a los contenidos poseer un rol causal. Sin embargo, también hemos visto en el apartado 1.2, dos argumentos bien fundados por los cuales las APs de los modelos computacionales de la mente si bien cumplen (1), no pueden cumplir con (2), y esto es así dado que la semántica de los símbolos *qua* semántica no tiene un rol causal porque sólo las propiedades sintácticas o formales son causalmente eficaces. La defensa de Fodor consiste en argumentar que mientras haya leyes intencionales causales que subsuman propiedades intencional es, entonces las propiedades intencional es tendrán un papel causal *qua* intencional es<sup>14</sup>.

Según Fodor, las leyes intencionales de la psicología científica son leyes *ceteris paribus* (al igual que las leyes de todas las ciencias especiales) que admiten excepciones, a diferencia de las leyes estrictas de la física que no tienen excepciones. En palabras de Fodor:

Las generalizaciones típicamente intencionales podrían ser de la forma: “Si ud. quiere —, y cree que no puede — a menos que haga — entonces, *ceteris paribus*, ud. intentará hacer —”. Por ejemplo, si ud. quiere *hacer un omelette*, y cree que no puede hacerlo a menos que *rompa algunos huevos*, entonces, a igualdad de condiciones, intentará hacer una acción que sea *la rotura de los huevos* (1994, p. 4),

<sup>14</sup> Esta idea se desarrolla en extensión en Fodor (1989), (1991) y (1994).

Otra manera de expresar la forma de estas leyes es: “los eventos F’s causan eventos G’s *ceteris paribus*”. Fodor sostiene que las afirmaciones singulares causales tienen que estar subsumidas por leyes causales. Así, el modelo causal al que se apela es el de la subsumición bajo leyes causales. Y lo que se sostiene es que la ocurrencia de F es nomológicamente suficiente para la ocurrencia de G. Ahora bien, las leyes intencionales están implementadas en mecanismos computacionales.

Un mecanismo es implementacional cuando en virtud de sus operaciones, la satisfacción del antecedente de una ley ocasiona de manera fiable la satisfacción de su consecuente [...]. Típicamente, aunque no invariablemente, los mecanismos que implementan las leyes de las ciencias están especificados en el vocabulario de alguna otra ciencia de nivel inferior (Fodor, 1994, p. 8).

Según esta visión, todas las ciencias menos la física, que es la ciencia básica, son ciencias especiales y en ese sentido tanto las leyes de la psicología como las de la biología o cualquier otra ciencia especial son leyes *ceteris paribus* y tendrían que estar implementadas, en última instancia, en mecanismos físicos. Pero teniendo en cuenta una visión estratificada de la realidad (en la cual la ciencia básica es la física y luego en un nivel superior, por ejemplo, vendría la biología y así hasta el nivel social), sólo hace falta mostrar cómo las leyes psicológicas se implementan en el nivel inferior.

Una teoría de la implementación responde a la pregunta: ¿Cómo los F’s causan los G’s?, y la respuesta consiste en especificar un mecanismo (en este caso, computacional) en el que la instanciación de F es suficiente para la instanciación de G. La figura que expresa esta relación es la siguiente (las flechas dobles indican que el antecedente es condición *causalmente* suficiente para el consecuente y las flechas simples indican la relación de condición suficiente para la *instanciación*)<sup>15</sup>:



<sup>15</sup> Según Fodor (1994, p. 10), por razones de simplicidad el diagrama supone que  $F \Rightarrow G$  no está múltiplemente realizado. Los distintos tipos de flecha son un agregado a la figura.

El hecho de que los MFs causan los MGs en el nivel computacional explica el hecho de que los Fs causen los Gs en el nivel intencional. Pero, por supuesto, se necesita una explicación de por qué F es condición suficiente para instanciar MF, y MG es condición suficiente para dar lugar a G<sup>16</sup>. La respuesta a esta cuestión de coinstanciación de propiedades computacionales e intencionales, en el fondo sigue siendo la misma respuesta que relacionaba a la sintaxis con la semántica de los símbolos. Antes vimos que la sintaxis “respeta” la semántica. Aquí, “respetar” significa que la relación entre el mecanismo computacional y los contenidos semánticos es “metafísicamente contingente”, su coinstanciación depende de hechos acerca del mundo. Esto no significa decir que es accidental sino que los hechos se obtienen tanto en el mundo efectivo como en todos aquellos mundos donde las leyes psicológicas se implementan de la misma forma que en el mundo efectivo.

Ahora bien, más allá de que habría dificultades para dar cuenta de la coinstanciación de las propiedades computacionales e intencionales<sup>17</sup>, y que la explicación del paralelismo entre la semántica y la sintaxis radica en decir simplemente “nuestro mundo es así”, que es lo mismo que decir que es un hecho bruto de la naturaleza cuando más bien las leyes de implementación requieren explicación, la postulación de leyes *ceteris paribus* ha recibido varias objeciones.

### 1.3.2 Las objeciones a las leyes *ceteris paribus*

Schiffer (1991) duda de que haya leyes *ceteris paribus*. Todo aquel que sostenga que puede haber este tipo de leyes tiene que hacer dos cosas. En primer lugar, tiene que explicar cuál es la naturaleza de la proposición expresada por “Los Fs causan los Gs, a igualdad de condiciones”. Y en segundo lugar, mostrar que la verdad de

<sup>16</sup> Cf. Fodor (1994), p. 10,

<sup>17</sup> Fodor (1994) intenta dar una solución a los casos Frege y los casos Gemelos que serían contraejemplos para esta coinstanciación. Los casos Frege muestran que la identidad de estados intencionales no es suficiente para la identidad de la implementación (porque más de un mecanismo implementacional puede corresponderse con el mismo estado intencional) y los casos Gemelos muestran que la identidad de implementación no puede ser suficiente para la identidad del estado intencional (porque más de un estado intencional puede corresponderse con el mismo mecanismo implementacional).

las proposiciones expresadas por las oraciones *ceteris paribus* determinan leyes (porque no cualquier oración verdadera determina una ley) y que además sean psicológicas. Con respecto a lo primero, Schiffer acepta que las oraciones *ceteris paribus* expresan proposiciones verdaderas pero, con respecto a lo segundo, éstas no constituyen leyes psicológicas. No son psicológicas porque las leyes que requieren las oraciones verdaderas *ceteris paribus* están expresadas en términos no intencional es y no son leyes porque no pueden usarse como premisas de una explicación por cobertura legal<sup>18</sup>.

Pérez (1995) muestra que la propuesta de Fodor se enfrenta a un dilema, o bien las leyes intencionales causales *ceteris paribus* son reducibles y en ese sentido no son intencional es, o bien no son ni leyes ni causales. Tomando en cuenta los desarrollos de Fodor (1991), Pérez distingue 4 tipos de leyes en función de un doble criterio, el de las excepciones y el de la reducción de leyes no-básicas a leyes básicas (2-4 serían las leyes de las ciencias no básicas): 1-leyes estrictas sin excepciones (las de la ciencia básica que es la física), 2-leyes que involucran esencias escondidas (la propiedad que figura en la ley tiene una sola realización posible, por ej. “agua” que sólo se realiza en “H<sub>2</sub>O”), 3-leyes *ceteris paribus* con meras excepciones (el realizador R1 de un estado mental M tiene un completador K1 pero éste puede no instanciarse, por eso sería una mera excepción a la ley “Los Fs causan Gs, *ceteris paribus*”)<sup>19</sup>, y 4-leyes *ceteris paribus* con excepciones absolutas (R1 no tiene K1, por ende, todos los casos de M que son realizados por R1 son excepciones absolutas a la ley *ceteris paribus* mencionada antes). Con respecto a las leyes *ceteris paribus* que son las que nos interesan, o bien son del tipo 3 o bien son del tipo 4. Si son del tipo 3, éstas admiten re-

<sup>18</sup> Schiffer (1991) llega a la idea de que el enunciado *ceteris paribus* se despliega de la siguiente manera: “Ms causan Bs sii para cada una de las “suficientemente muchas” realizaciones D de M hay una condición del mismo nivel C tal que D-&-C es causalmente suficiente de una manera no superflua para un evento B” (p, 7). Este bicondicional da el significado de las oraciones *ceteris paribus*. Dado que hay instancias de sustitución verdaderas del lado derecho del bicondicional, las oraciones *ceteris paribus* expresan proposiciones verdaderas. Pero las leyes aludidas por estas oraciones *ceteris paribus* dicen que una condición no-física D-&-C es nomológicamente suficiente para la ocurrencia de B. Por otro lado, ninguna instancia de sustitución del lado derecho del bicondicional es una ley de ningún tipo.

<sup>19</sup> Dada la múltiple realizabilidad de una propiedad mental, la propiedad física que la realiza puede no ser suficiente para causar G. Los completadores son propiedades o circunstancias físicas que hay que adicionar al realizador de un estado mental para que el estado mental sea suficiente para causar G.

ducciones locales de tipos (identidad tipo-tipo en una misma especie o conjunto de individuos), así “Los F’s causan Gs, *ceteris paribus*” podría reducirse a “Los R1 + K1 causan Gs, *ceteris paribus*”; “Los R2 + K2 causan Gs, *ceteris paribus*”;...; “Los Rn + Kn causan Gs, *ceteris paribus*”, con lo cual las leyes intencional es podrían reducirse a leyes no-intencionales y en ese sentido no se ve por qué la psicología requeriría de leyes específicas. Si son del tipo 4, no serían reducibles porque hay Rs que no causan Gs, pero la ley no apoyaría subjuntivos del tipo “Si esto fuera un F causaría G”<sup>20</sup>, y además no serían enunciados causales porque no habría conjunción constante entre F y G, con lo cual los enunciados *ceteris paribus* no serían ni legaliformes ni causales.

Si objeciones como las mencionadas resultan ser válidas, entonces si no puede haber leyes intencionales causales, las propiedades intencionales no tendrían poder causal. Ahora bien, del hecho de que no haya leyes intencional es no se sigue que las propiedades intencionales sean epifenoménicas porque que haya leyes intencionales causales es sólo condición suficiente (y no necesaria) para que las propiedades intencionales tengan poder causal<sup>21</sup>. Pero lo que sí se sigue es que si no hay leyes *ceteris paribus*, la TRCM se queda sin una explicación del rol causal de las propiedades semánticas/intencionales. Y eso alcanza como evidencia para lo que vengo sosteniendo, esto es, la dificultad de compatibilizar (1) con (2).

## 2. Mejor contentarse con un realismo moderado

Si se acepta la idea de que en los dos tipos de modelos computacionales que compiten para dar cuenta de la arquitectura mental, el conexionismo y la arquitectura clásica, se cumple (1) pero no (2)<sup>22</sup>, ¿se desprende que el realismo acerca de las APs es inviable? No ne-

<sup>20</sup> Una concepción bastante consensuada acerca de las leyes consiste en que éstas son generalizaciones legaliformes que son apoyadas por subjuntivos y contrafácticos.

<sup>21</sup> Este punto es de Pérez (1995). Se han propuesto otras nociones para dar cuenta de la relevancia causal de las propiedades intencionales pero también tienen problemas, Véase la concepción contrafáctica de Lepore y Loewer (1987) y la respuesta de Fodor (1991).

<sup>22</sup> Por supuesto, se podría cuestionar la idea misma de intentar explicar la intencionalidad a partir de modelos computacionales de la mente. Sin embargo, el objetivo de este trabajo es una crítica interna y no externa.

cesariamente, se puede sostener un realismo sin causalidad pero realismo al fin. Y, ¿se desprende del hecho de que no haya leyes intencionales que no se puede tener una psicología científica? No necesariamente, se puede obtener una psicología científica aunque no haga uso del modelo de subsunción legal-causal, pero científica al fin. En lo que sigue intentaré una defensa de estas respuestas.

## 2.1 El realismo moderado

En este apartado voy a sostener que de las tesis (i) que los argumentos eliminativistas que concluyen con buenas razones que las APs no tienen rol causal en los modelos computacionales de la mente (apartado 1.2) y (ii) que la propuesta de leyes intencional es causales para otorgar rol causal a las propiedades intencionales no funciona (apartado 1.3), no se sigue (iii) ni el eliminativismo fuerte ni el eliminativismo débil. Comenzaré con la cuestión del eliminativismo fuerte.

En primer lugar, quisiera hacer una observación muy obvia y es que si resulta que los contenidos intencionales no tienen papel causal *qua* intencional es (supongamos que ninguna propuesta de condiciones necesarias y/o suficientes para que las propiedades intencionales tuvieran rol causal funcionara), de aquí no se sigue que los estados intencionales no existan. Afirmar que los estados intencionales son epifenómenos consiste en sostener que son subproductos de estados físicos y en ese sentido son causados por estados físicos pero ellos no causan nada *qua* intencionales (ni otros estados físicos ni mentales). De esto no se sigue que aquello causado por los estados físicos no existe. Así, del hecho de que las APs no tengan la propiedad de “ser causales” no se sigue el eliminativismo fuerte<sup>23</sup>.

En segundo lugar, se podría pensar que las APs existen porque aun cuando las entidades postuladas por los modelos computacionales no tengan la propiedad de “ser causal es”, tienen otras propiedades como la de “ser semánticamente interpretables” (y “funcio-

<sup>23</sup> Alguien podría pensar que para los filósofos que siguen el dictum de Alexander (1920): “ser real es tener poderes causales”, los estados epifenómicos no serían reales. Sin embargo, tal como afirma Kim (1996, cap. 6), el dictum no constituye un argumento en contra del epifenomenalismo porque éste le otorga a lo mental un lugar en la estructura causal del mundo (en tanto que “es causado”, lo mental no está en completo aislamiento del resto del mundo), por lo tanto el dictum no compromete la realidad de los estados mentales epifenómicos.

nalmente discretas” y cualquier otra que se piense que tienen). Ahora bien, lo que quiero defender es que aun cuando las entidades postuladas por los modelos computacionales de la mente no tengan ninguna de las propiedades que se suelen atribuir a las creencias (o APs en general), esto no implica que las creencias no existan. Es decir, no hay argumento *a priori* que permita pasar del eliminativismo débil (entendido en una de las acepciones mencionadas como la tesis de que las “creencias” no figurarán en la ontología de las ciencias) al eliminativismo fuerte (las creencias no existen).

Este es el camino que toma Stich, unos años después del artículo de 1990. Allí sostenía que si la PSC resulta falsa entonces, se sigue el eliminativismo de las APs. Stich y Warfield (1995) atemperan el eliminativismo desdoblado el condicional en dos partes: (i) si ciertas hipótesis conexionistas resultan verdaderas, entonces la PSC está seriamente equivocada, y (ii) si la PSC está seriamente equivocada, entonces se sigue el eliminativismo de las APs. Lo que intentan mostrar es que (ii) no se sigue de (i) porque haría falta alguna premisa adicional muy problemática para sostener que del hecho de que la PSC esté seriamente equivocada se concluya el eliminativismo de las APs. Lo que se modifica del argumento de 1990 es la idea de que no se puede concluir de la verdad de la tesis eliminativista débil, la verdad de la tesis eliminativista fuerte (sin suponer premisas problemáticas).

Stich y Warfield (1995) sostienen que aun cuando las entidades postuladas por el conexionismo no tuvieran ninguna de las propiedades (a)-(c), no se puede concluir el eliminativismo sin invocar premisas problemáticas. Para que del eliminativismo débil se siga el fuerte, un eliminativista tendría que defender o bien una teoría descriptiva de la referencia o bien la idea de que ciertos conceptos tienen propiedades constitutivas. La teoría descriptiva sostiene que la referencia de un concepto está determinada por aquellas entidades que satisfacen una cierta descripción del concepto. Pero esto tiene dificultades porque vuelve al eliminativismo trivialmente verdadero (si pequeñas diferencias en la descripción determinan referentes diferentes) o trivialmente falso (si grandes diferencias no son tomadas en cuenta para determinar el referente). La concepción de las propiedades constitutivas sostiene que hay ciertas propiedades que son conceptualmente necesarias a la manera en que “ser no casado” sería constitutivo del concepto “soltero”. Así, si se piensa que la modularidad proposicional ((a)-(c)) es una propiedad constitutiva de las APs, el argumento eliminativista toma la si-

guiente forma: si la ciencia muestra que no hay nada que tenga esa propiedad, las APs no existen. El problema que tiene esta concepción es cómo establecer que cierta propiedad resulta constitutiva de un concepto, parece que la única manera sería defender la distinción analítico-sintética y es de público conocimiento filosófico el ataque quineano a esta distinción<sup>24</sup>.

Acuerdo con la idea de que aun cuando los estados conexionistas o de la arquitectura clásica interpretados como creencias no tuvieran ninguna de las propiedades de las APs haría falta algo más que lleve a la eliminación de las APs. Pero, a diferencia de Stich y Warfield (1995), creo que la cuestión del eliminativismo no es un asunto de alguna premisa que falta en el argumento eliminativista. En otra oportunidad<sup>25</sup> señalé que aun utilizando premisas problemáticas como las que se sugieren en Stich y Warfield (1995) (basadas en una teoría de la referencia en particular o en propiedades constitutivas de los conceptos intencional es), esto no permitiría dirimir la cuestión porque la eliminación de entidades es una cuestión histórica. Los cambios de teoría ontológicamente radicales no dependen de argumentos *a priori* sino de las prácticas científicas que están imbuidas en contextos sociales más amplios.

Por la misma época, Stich vuelve a debilitar un poco más su argumento eliminativista y en un artículo de 1996 termina afirmando esto mismo. Para él la cuestión del eliminativismo fuerte se establece en general por factores sociales y políticos internos a la disciplina en cuestión, y en el caso particular de los estados intencionales habría también factores sociales y políticos más amplios, ya que la existencia o no de creencias y deseos afectaría a una gama amplia de ámbitos humanos (por ejemplo, la decisión acerca de cómo tratar a personas con problemas mentales o emocionales). A continuación me ocuparé de la cuestión del eliminativismo débil.

Hasta aquí estoy de acuerdo con la idea de que no hay un argumento conceptual que permita el paso del eliminativismo débil al fuerte. Sin embargo, en toda esta polémica nunca se cuestionó la posibilidad del eliminativismo débil. Stich y sus colaboradores consienten en la idea de que si ciertos modelos de la arquitectura mental resultan verdaderos, entonces las APs, si bien existen, no formarán parte de la psicología científica. Lo que quiero cuestionar ahora es la afirmación de que la noción de estados con descripción semán-

<sup>24</sup> Cf. Quine (1951),

<sup>25</sup> Cf. Skidelsky (1996).

tica no forme parte de las teorías computacionales de la mente. Mi idea es que el eliminativismo débil tampoco se sostiene' porque las APs tienen un rol explicativo en estas teorías, es más, si no lo tuvieran estas teorías no serían cognitivas y en tanto que tienen un rol explicativo formarán parte de la psicología científica. Así, lo que intento atacar es una de las formas que adopta el eliminativismo débil y que mencioné antes que consiste en la tesis de que los estados caracterizados intencionalmente no jugarán un rol explicativo en la psicología científica. Si logro esto, entonces, la otra forma que adopta el eliminativismo débil (que las APs no formarán parte de la psicología científica) no se sostiene<sup>26</sup>.

Es ampliamente aceptada en el ámbito psicológico y filosófico la distinción de niveles de explicación de una teoría cognitiva que trazó D. Marr (1982, 1977), quien distingue tres niveles jerárquicos de explicación desde los cuales entender la/s tarea/s que lleva a cabo cualquier sistema de procesamiento de información. El nivel superior es el *nivel computacional* que consiste en un análisis abstracto (formal) del problema que va a ser computado. Se relaciona con la pregunta de qué va a ser computado y por qué. El nivel siguiente es el *nivel algorítmico* en el que se especifica la representación para el input y el output y el algoritmo para la transformación. Y el nivel inferior es el *nivel del hardware* en el que se especifica cómo puede realizarse físicamente la representación y el algoritmo.

Si bien el nivel computacional tal como lo definió Marr es una caracterización matemática de la función computada por el sistema, con posterioridad se lo ha redefinido de manera de incluir en este nivel una descripción en términos semánticos de la capacidad cognitiva que realiza el sistema<sup>27</sup>. Es necesario un nivel semántico para dar cuenta de lo que hace el sistema porque de otra manera no habría una caracterización cognitiva del mismo y por ende, no podría considerárselo un sistema cognitivo. Esto se relaciona con lo que se dijo antes acerca de la diferencia entre una teoría computacional y

<sup>26</sup> Un argumento a favor del eliminativismo débil basado en la ineficacia causal del contenido sería el siguiente: Si el contenido no tiene eficacia causal, entonces nada ocurre *en virtud* del contenido, y si nada ocurre *en virtud* del contenido, entonces apelar al contenido no tiene ningún rol explicativo genuino en la psicología madura. Mi intención es mostrar que no se requiere apelar a la eficacia causal del contenido para que éste tenga un rol explicativo genuino.

<sup>27</sup> Véase como un ejemplo Sterelny (1990), quien rebautiza los niveles con las expresiones "nivel ecológico", "nivel computacional" y "nivel de implementación física", respectivamente.

una teoría cognitiva. La teoría computacional da una descripción formal, es decir en términos no semánticos/intencionales, en cambio, la teoría cognitiva da una descripción semántica que depende del entorno en donde esté inserto el sistema<sup>28</sup>. Mientras que las computaciones de una función matemática son independientes del entorno, el contenido semántico de los estados del sistema computacional va a estar determinado por el entorno. Así, una misma función matemática puede tener distintas interpretaciones semánticas según el dominio representado. El contenido tiene un rol explicativo porque permite conectar las caracterizaciones no semánticas/intencionales del sistema con el entorno del mismo, de manera que es posible describir los estados internos de un sistema en términos semánticos<sup>29</sup>.

Lo interesante de la postulación de modelos computacionales de la arquitectura mental descansa en la idea de que hay un nivel de procesamiento de la información gracias al cual se pueden realizar generalizaciones acerca de los sistemas que ejecutan cierta capacidad cognitiva. Si este nivel semántico de descripción de lo que hacen los sistemas cognitivos no existiera, entonces ahí sí nos quedaríamos sin una psicología porque las únicas explicaciones disponibles serían en términos físicos (no semánticos/intencionales). El papel explicativo de los contenidos mentales permite entonces explicar las capacidades cognitivas sin necesidad de apelar al papel causal de dichos contenidos<sup>30</sup>.

<sup>28</sup> Aquí, “teoría computacional” se refiere al nivel algorítmico de Marr y “teoría cognitiva” al nivel computacional/semántico.

<sup>29</sup> Conuerdo con Egan (1995, p. 189) quien afirma que “es sólo bajo la interpretación de algunos de los estados del sistema como representaciones de propiedades distales (como profundidad, o la reflectancia de las superficies) que los procesos que tienen una caracterización matemática dada por una teoría computacional se revelan como *visión*”.

<sup>30</sup> Mi intención fue mostrar, en relación con el eliminativismo fuerte, que no hay manera de argumentar *a priori* (del hecho de que los estados intencionales no tengan rol causal o, en general, ninguna de las propiedades que hemos visto) en favor de la inexistencia de las APs. Mi estrategia del *realismo* moderado no se basa en argumentar positivamente en favor del realismo porque hacerlo sería aceptar el desafío escéptico y no es una línea que me parezca adecuada seguir. En cambio, en relación con el eliminativismo débil, argumento positivamente en favor de un realismo *moderado* sosteniendo que las creencias (si bien no tendrían un rol causal) tienen un rol explicativo. No se supone que esto actúe como un argumento en favor de lo primero (esto es, del realismo). Sin embargo, si alguien quisiera conectar en psicología ‘realismo’ con ‘rol explicativo’ y sostener que los estados intencionales son reales en tanto tienen un rol explicativo, mis simpatías irían con él. Porque dado que el rol expli-

Ahora bien, por un lado parece que para que una psicología científica tenga (1) estados descriptos en términos semánticos, no requiere que (2) el contenido semántico tenga rol causal. Pero, habíamos visto que para el RIA la posibilidad de una psicología científica descansa en el papel causal de los contenidos de esos estados que a su vez depende de si son posibles las leyes intencionales, y dado que aceptamos que el estatus de esas leyes era dudoso, ¿esto significa que si no hay leyes intencionales, no puede haber una psicología científica?

## 2.2 La explicación psicológica

Los argumentos desarrollados en 1.2 y 1.3 concluyen que (i) las APs no tienen rol causal en los modelos computacionales de la mente y que (ii) es muy dudoso que haya leyes intencionales causales, pero de esto (iii) no se sigue que no se pueda obtener una psicología científica. En consonancia con la crítica a las leyes causales intencionales, Schiffer (1991, p. 16) afirma lo siguiente:

Cuando yo leo [textos de] biología, me cuesta mucho encontrar algo que parezca una explicación que invoque leyes, y creo que sé por qué. Supongamos que Ud. inventó una ratonera que se activa con un resorte y tuvo que explicar *cómo funcionaba*. Ud. diría que, cuando la máquina funciona, es porque un ratón mordisquea el queso puesto en un mecanismo disparador; el movimiento causado por el mordisqueo dispara una reja acoplada a un resorte estirado; etc. Pero no mencionaría ninguna ley. Quizá si continuara la explicación con una cadena causal lo suficientemente larga, llegaría a leyes; pero serían leyes de la física, no leyes acerca de la teoría de las ratoneras. De la misma manera, la mayor parte de la biología se ocupa de explicar cómo funcionan varios mecanismos –piense en la explicación de la fotosíntesis– y parece que esas explicaciones no invocan *leyes* biológicas, estrictas o *ceteris paribus*.

---

cativo de los estados intencionales es conectar las caracterizaciones no-semánticas/intencionales de un sistema cognitivo con el entorno del mismo (y esto sólo es posible en términos semánticos/intencionales, o sea logrando (1)), y dado que si no hay un nivel de descripción intencional en una teoría psicológica, la teoría no sería psicológica (sino neurofisiológica o el nivel no-intencional que se desee), no habría el peligro de que en una psicología futura los estados intencionales no tuvieran este rol explicativo y con ello no fueran reales, porque si éste fuera el caso, la teoría ya no sería una teoría *psicológica*.

La cita se refiere a la biología pero es probable que la mayoría de las ciencias especiales, incluida por supuesto la psicología, se ocupen de explicar cómo funciona un mecanismo y en ese sentido no requieren apelar a leyes especiales. El nivel donde hay que buscar las leyes es el nivel físico, por ende, las ciencias especiales no tendrán leyes propias, y por ende, éstas no serán ni intencionales ni causales, en todo caso las explicaciones de los mecanismos cognitivos apelarán a leyes causales de la física.

Los que creen que las ciencias especiales deben tener leyes especiales parten básicamente de la motivación de que las propiedades intencionales sólo pueden tener un rol genuinamente explicativo si aparecen en explicaciones intencionales genuinas y éstas sólo pueden adoptar la forma de las explicaciones del modelo de cobertura legal-causal. Una alternativa al modelo de subsunción bajo leyes causales es la “explicación funcional” propuesta por Cummins (1975, 1983). La idea es bastante antigua, pero se ve que ciertos filósofos (como es el caso de Fodor) siguen apegados al más antiguo modelo nomológico deductivo de Hempel y Oppenheim (1948). No me voy a extender mucho en los detalles de esta propuesta porque lo único que me interesa mostrar es que es viable otro modelo explicativo.

Cummins propone que las explicaciones psicológicas se ocupan de explicar ciertas propiedades de un sistema y en este sentido son “teorías de propiedades”. El objetivo de una “teoría de propiedades” es responder la pregunta ¿en qué consiste para el sistema S tener la propiedad P (o más precisamente, instanciar la propiedad P)? Para ello se apela a las propiedades de los componentes de S y cómo están organizadas. La idea principal es explicar una capacidad cognitiva compleja descomponiendo el sistema en subcapacidades más simples y menos sofisticadas y problemáticas que la capacidad que se quiere explicar tal que la manifestación de las capacidades analiza doras (especificables en un programa o un diagrama de flujo) equivale a la manifestación de la capacidad analizada<sup>31</sup>.

Las capacidades de procesamiento de la información son propiedades disposicionales complejas y su análisis se denomina ‘análisis funcional’ (puesto que la capacidad que se analiza es una función del sistema). El primer paso del análisis funcional consiste en el *análisis* de la capacidad cognitiva P y el segundo en dar cuenta de la *instanciación* en S del análisis de P. De manera que si S posee los

<sup>31</sup> De hecho, si uno pasa revista a la literatura de la ciencia cognitiva es justamente este tipo de explicación la que prepondera.

componentes organizados de la manera especificada en el análisis, entonces S posee P. Así, la explicación funcional concluye cuando el sistema S ejecuta el programa especificado por el análisis (S tiene que tener la estructura necesaria para cumplir las instrucciones del programa), de manera que todo el peso de la explicación funcional radica, en última instancia, en la especificación de las capacidades analizadoras en el programa.

La teoría de propiedades no requiere leyes causales como sí las requieren las “teorías de la transición” (que son teorías de subsunción bajo leyes causales) que intentan responder ¿por qué el sistema cambia del estado S1 al S2? En este caso se explican los cambios como efectos de causas previas y eso requiere apelar a leyes causales. En cambio, en la “teoría de propiedades” se utilizan leyes (de instanciación y composición)<sup>32</sup> que no son causales, esto es, no se explica la posesión de P como efecto de alguna causa. El punto central es que los fenómenos psicológicos no se explican subsumiéndolos bajo leyes causales sino dando cuenta de ellos como manifestaciones de capacidades explicadas por análisis<sup>33-34</sup>.

<sup>32</sup> Una ley de instanciación (“Cualquier cosa que tenga los componentes C1...Cn organizados de manera  $\phi$  (teniendo un análisis [C1...Cn, O]) tiene la propiedad P”, Cummins, 1983, p. 17) es una afirmación legaliforme que especifica cómo una propiedad P se instancia en un tipo especificado de sistema S, Una ley de composición (“S tiene el análisis [C1...Cn,O]”, Cummins, 1983, p. 17) es una afirmación legaliforme que especifica el análisis de un tipo de sistema, (por ejemplo, el modelo de la doble hélice del ADN o la composición molecular del agua). Las leyes de instanciación no son causal es porque para que lo sean debe aparecer el par causa-efecto en sus instancias, Pero, requieren de explicación, esto es, se derivan de “atribuciones nómicas” que son afirmaciones legaliformes que expresan que todo X tiene cierta propiedad P (por ejemplo, la ley de inercia y de la gravedad).

<sup>33</sup> Cuando la propiedad disposicional P se manifiesta, hay una transición de estados en el sistema, por ende hay una teoría de la transición que especifica las leyes de la transición de la disposición P y de las disposiciones que componen la disposición P, en este sentido una teoría de la transición especifica el *explananda* para una teoría de propiedades, no dice qué es para P tener la propiedad disposicional p, Cf. Cummins, 1983, cap. 1.

<sup>34</sup> Dado que la explicación nomológica deductiva [N-D] no requiere leyes causales sino tan sólo leyes, se podría pensar que se le puede dar una forma N-D a la explicación funcional. Sin embargo, hay consideraciones generales y particulares para pensar que las leyes a las que apela la explicación N-D son causales. Por un lado, la explicación N-D es una expresión de la comprensión empirista de la causalidad y la explicación causal (véase Collins, 1987, cap. 7). Por otro lado, según Cummins (1983, cap. 1) las teorías de la transición subsumen eventos particulares bajo leyes (de ahí que son explicaciones N-D), estas leyes tienen que ser causales (subsumen pares causa-efecto) porque si no, estas teorías no serían explicativas (dado que su objeto es explicar cambios como efectos), de ahí que las leyes de las explicaciones

Ahora bien, se podría pensar que la propiedad de “ser intencional” de los sistemas cognitivos podría explicarse apelando a subcapacidades no intencionales en conexión con ciertas propiedades del entorno que determinarían el contenido de los estados intencionales. El desarrollo de este punto llevaría más allá del objetivo de este trabajo, es suficiente con haber mostrado que existe una alternativa de explicación para la psicología científica que concuerda perfectamente con lo que hemos visto en 2.1 con respecto al rol explicativo de los estados intencionales. De manera que no sería cierto que si no hay leyes causales intencionales no habría psicología científica.

La explicación funcional está en consonancia con la idea presentada anteriormente de que no hace falta leyes *especiales* para la explicación psicológica. Sin embargo, se diferencia de la anterior porque ésta apela en última instancia al nivel físico (leyes de la física), mientras que la explicación funcional está pensada para un nivel más abstracto que es el computacional (leyes computacionales). Si de lo que se trata es de dar cuenta de por qué un mecanismo tiene cierta capacidad, entonces no hace falta apelar a leyes causales en el nivel computacional; sin embargo, si de lo que se trata es de explicar la transición de estados en un sistema computacional, entonces ahí sí se baja al nivel físico para dar cuenta del paso de la causa al efecto. Cummins cree que lo que convierte en explicativa a una explicación psicológica no es esto último (una “teoría de la transición”), sino lo primero (una “teoría de propiedades”). Así, tanto la propuesta de Schiffer como la de Cummins acuerdan en que la explicación psicológica no requiere de leyes causales especiales.

### 3. Conclusión

A lo largo de este trabajo intenté mostrar que el RIA quiere demasiado al intentar conciliar estados (1) con contenido semánticamente interpretable y (2) que tengan rol causal en la producción de la conducta, y que si (2) no se puede lograr, esto no significa, por un lado, ni un eliminativismo débil (porque las APs tendrían un rol explicativo dado que conectarían las caracterizaciones computacio-

---

N-D para fenómenos psicológicos sean causales. De manera que transformar una explicación funcional en una N-D oscurecería la diferencia entre explicar cambios y explicar propiedades.

nales con el entorno del sujeto y en ese sentido formarán parte de una psicología científica) ni uno fuerte (porque no hay argumento conceptual que permita establecerlo), ni significa, por otro lado, la no obtención de una psicología científica (porque la psicología puede hacer uso de otro modelo explicativo que no consista en la subsunción de los fenómenos psicológicos bajo leyes causales intencionales). Contentarse con menos (quedarse sólo con (1)) sigue siendo una forma de realismo, moderado, pero realismo al fin y que es perfectamente compatible con la idea de obtener una psicología científica.

## BIBLIOGRAFIA

- Alexander, S. (1920), *Space, Time, and Deity*, London, Macmillan.
- Collins, A. (1987), *The Nature of Mental Things*, Notre Dame, Indiana, University of Notre Dame Press.
- Cummins, R. (1975), "Functional Analysis", *Journal of Philosophy* 72 (20): 741-64.
- (1983), *The Nature of Psychological Explanation*, Cambridge, Mass., MIT Press.
- (1989), *Meaning and Mental Representation*, Cambridge, Mass., MIT Press.
- Dennett, D. (1991), "Real Patterns", en (1998), *Brainchildren*, Cambridge, Mass., MIT Press, pp. 95-120.
- Egan, F. (1995), "Computation & Content", *The Philosophical Review*, vol. 104, 2: 181-203
- Fodor, J. (1980) "Methodological Solipsism Considered as Research Strategy in Cognitive Psychology", en: (1981), *Representations*, Cambridge, Mass., MIT Press, pp. 225-253.
- (1987), *Psicosemántica*, Madrid, Tecnos, 1994.
- (1989), "Making Mind Matter More", *Philosophical Topics*, 17,1: 59-79. Reimpreso en Fodor, J. (1990), *A Theory of Content and Others Essays*, Cambridge, Mass., MIT Press.
- (1991), "You Can Fool Some of The People All of the Time, Everything Else Being Equal, Hedged Laws and Psychological Explanations", *Mind*, vol. 100: 19-34.
- (1994), *The Elm and the Expert*, Cambridge, Mass., MIT Press.
- (1999), *Conceptos. Donde la ciencia cognitiva se equivocó*, Barcelona, Gedisa.
- Hempel, C. y Oppenheim, P. (1948), "Studies in the Logic of Explanation", *Philosophy of Science*, 15: 135-175.

- Kim, J. (1994), "Supervenience", en: S. Guttenplan (ed.) (1994), *A Companion to the Philosophy of Mind*, Oxford, Blackwell.
- (1996), *Philosophy of Mind*, Oxford, Westview.
- Lepore, E. y Loewer, B. (1987), "Mind Matters", *Journal of Philosophy*, 84, 11: 630-42.
- Marr, D. (1977), "Artificial Intelligence-A personal view", en: J. Haugeland (ed.) (1981), *Mind Design*, Cambridge, Mass., MIT Press, pp. 129-142.
- (1982), *Vision*, San Francisco, Freeman.
- Pérez, D. (1995), "¿Leyes causales intencionales? Sobre los argumentos de Fodor a favor de las leyes intencionales", *Revista Latinoamericana de Filosofía*, vol. XX, 1, n°1: 67-81.
- Quine, W. (1951), "Two dogmas of empiricism", en: (1953), *From a Logical Point of View*, New York, Harper and Row.
- Ramsey, W., Stich, S. y Garon, J. (1990), "Connectionism, Eliminativism and the Future of Folk Psychology", en W. Ramsey, S. Stich y D. Rumelhart (eds.) (1991), *Philosophy and Connectionist Theory*, Hillsdale, NJ, LEA, pp. 199-228.
- Schiffer, S. (1991), "Ceteris Paribus Laws", *Mind*, vol. 100: 1-17.
- Skidelsky, L. (1996), "Creencias: 'Folk Psychology' y modelos computacionales de la mente", ponencia presentada en el *Coloquio de Filosofía Teórica*, Sadaf, Argentina, 11 y 12 de octubre.
- Smolensky, P. (1995), "On the Projectable Predicates of Connectionist Psychology: A Case for Belief", en: C. Macdonald y G. Macdonald (eds.), *Connectionism: Debates on Psychological Explanation*, Cambridge, Basil Blackwell, pp. 357-394.
- Sterelny, K. (1990), *The Representational Theory of Mind*, Oxford, Basil Blackwell.
- Stich, S. (1992), "What is a Theory of Mental Representation?", *Mind*, vol. 101: 243-261.
- (1996), "Deconstructing the Mind", en: *Deconstructing the Mind*, Oxford, Oxford University Press, pp. 3-90.
- Stich, S. y Warfield, T. (1995), "Replay to Clark and Smolensky: Do Connectionist Minds Have Beliefs?", en: C. Macdonald y G. Macdonald (eds.), *Connectionism: Debates on Psychological Explanation*, Cambridge, Basil Blackwell, pp. 395-411.