



FILO:UBA
Facultad de Filosofía y Letras
Universidad de Buenos Aires

P

Arquitectónica de la mente naturalizada

Autor:

Wajnerman Paz, Abel

Tutor:

Cresto, Eleonora

2014

Tesis presentada con el fin de cumplimentar con los requisitos finales para la obtención del título Doctor de la Facultad de Filosofía y Letras de la Universidad de Buenos Aires en Filosofía

Posgrado



FILO:UBA
Facultad de Filosofía y Letras

FILODIGITAL
Repositorio Institucional de la Facultad
de Filosofía y Letras, UBA

Tesis de doctorado:
Arquitectónica de la mente naturalizada

Directora: Eleonora Cresto

Consejero: Federico Penelas

Doctorando: Abel Wajnerman Paz

Índice

1. Introducción: Los dos desafíos para la naturalización de la mente.....	7
--	----------

Parte I

2. Una teoría informacional para los casos Frege.....	19
--	-----------

3. Contenido cognitivo y forma sintáctica.....	35
---	-----------

4. Representaciones perceptuales y sistemas perceptuales.....	59
--	-----------

5. La distinción y relación conceptual/perceptual y un anti-empirismo de conceptos.....	72
--	-----------

6. Conocimiento como información doxástica y un anti-empirismo epistémico.....	85
---	-----------

Parte II

7. Computación y procesamiento neuronal.....115

8. Cognición y computación.....128

9. Cognición y procesamiento neuronal.....152

10. Realización.....171

11. Conclusión: El sistema informacional de la mente.....197

Agradecimientos

Quiero agradecer ante todo a Eleonora Cresto. Además de lo que se espera comúnmente de una directora o director (y muy poco comúnmente llega), es decir, una insistente lectura atenta, crítica y constructiva que hace que nuestras no tan buenas ideas puedan ir tomando la forma de algo inteligible, he encontrado en ella una persona extraordinariamente atenta, gentil, honesta y comprensiva. Realmente estoy muy agradecido y contento de haber tenido la oportunidad de compartir este momento importante de la formación con tamaña filósofa y persona.

Quiero agradecer también especialmente a Liza Skidelsky, por haberle dado un piadoso golpe de gracia a encarnaciones previas de muchas ideas que aquí figuran y haberme ayudado *enormemente*, por medio de intensos momentos de catarsis conceptual, a mejorar mi perspectiva sobre casi todos los temas que aquí se tratan y sobre casi cualquier otro.

Además quiero agradecerle al grupo que dirige Skidelsky, compuesto por ella misma, Sergio Barberis, Sabrina Haimovici, Mariela Destefano, Nicolás Serrano y Fernanda Velazquez, por haberse convertido, muy a pesar de todos ellos, en mi primer *hogar* filosófico. Realmente han hecho para mí de la vida filosófica, de la tarea de discusión filosófica, al mismo tiempo una actividad extraordinariamente nutritiva y una actividad de un disfrute y calidez únicos.

Agradezco a mi mujer, Julieta Picasso, dueña de todo lo que alguna vez fue mi individualidad y autodeterminación, condición de posibilidad de todo lo bueno y lo malo, incluyendo lo bueno y lo malo de esta tesis.

Por último y por primero, agradezco a mis padres, Alejandro Wajnerman y Mercedes Paz, por no haber sido nunca espantados por mi entusiasmo filosófico sino, por el contrario, haber adoptado, cuidado y cultivado esta monstruosidad por tantos años, con tanto amor.

Capítulo I: Dos desafíos para la naturalización de la mente

“Bajo el gobierno de la razón, nuestros conocimientos no pueden ser, en general, una rapsodia, sino que deben constituir un sistema [...] Entiendo por sistema la unidad de los múltiples conocimientos bajo una idea.” Kant (1781/1787).

1. La teoría informacional del contenido mental

El propósito medular de este trabajo es ofrecer una nueva defensa de la psicología intencional sobre el supuesto de alguna forma de fisicalismo. Es decir, argumentaré que los estados mentales son intencionales mostrando que los estados intencionales son idénticos a estados físicos. Esta tesis ha sido famosa y recurrentemente defendidas por Jerry Fodor. Mi objetivo es profundizar dicha defensa refinando lo que considero que es la pieza fodoriana medular de la misma: la teoría informacional del contenido mental.

La primer idea que guiará este trabajo es entonces, alguna forma de fisicalismo. Como afirma Fodor (2008), la tesis de que todo lo que existe es, en última instancia, físico, es un supuesto que subyace a toda empresa científica (Fodor 2008), parte del núcleo duro de la actividad científica. Si queremos tener una comprensión filosófica de los fenómenos mentales que pueda ser *complementaria* con la perspectiva científica sobre ellos, parece que debemos llevar adelante nuestra tarea de elucidación con este supuesto en vista.

La tesis fisicalista se ha entendido de al menos dos modos diferentes. Durante mucho tiempo se la sostuvo como una tesis de identidad de *tipos* específicamente referida a tipos *psicológicos*. El fisicalismo ha sido simplemente la tesis de la identidad psico-física de tipos. Esta es la tesis de que cada tipo de estado o propiedad psicológica es idéntico a un tipo de estado o propiedad física del cerebro o el sistema nervioso (Smart 1959). Sin embargo, también puede entenderse, en la línea del párrafo

anterior, como un caso de la regla general de que todo tipo de propiedad es idéntico con algún tipo de propiedad física. Es sabido que primero Putnam (1967), y después Fodor y Block (1972), pusieron en duda este supuesto hasta ese entonces ampliamente aceptado sobre la relación entre los estados mentales y los estados físicos. Estos autores cuestionaron la tesis de la identidad de tipos mostrando o bien la implausibilidad de que diferentes casos de una propiedad psicológica, por ejemplo, sean casos de un mismo tipo físico o bien presentando situaciones en las que esto efectivamente no ocurre. A partir de casos como estos, la forma de fisicalismo más comúnmente aceptada es la tesis del fisicalismo de casos. La formulación clásica de Fodor afirma que todos los eventos particulares que caen bajo las leyes de las ciencias especiales (entre ellas, la psicología) son eventos particulares físicos (es decir, eventos que caen bajo leyes de la física) (Fodor 1974). Así, todo caso de un evento o propiedad psicológica es idéntico a un caso de un evento o propiedad física. Una formulación más reciente afirma que para todo particular (objeto, evento, o proceso) x efectivamente existente, existe algún particular físico y tal que $x=y$ (Stoljar 2009).

En la formulación de Stoljar se ve más claramente el matiz normativo del fisicalismo de casos. El hecho de que para cualquier instancia de un tipo de propiedad, exista una instancia de una propiedad física con la que sea idéntica, es condición necesaria para que podamos decir que existe. Sólo podemos afirmar que existe una propiedad P cualquiera postulada por alguna de las ciencias particulares si sus casos son casos de algún tipo físico. La formulación de Fodor, por el contrario, asume que las instancias de las propiedades psicológicas, por ejemplo, pueden ser individuadas en términos físicos. Sin embargo, no va de suyo que el fisicalismo de casos sea amigable con los estados mentales o psicológicos. En particular, no es obvio qué determina el fisicalismo sobre los estados psicológicos si entendemos, con Fodor (por ejemplo, Fodor 1994 y 2008) que los procesos psicológicos son procesos que involucran estados que deben ser individuados, al menos en parte, en virtud de sus contenidos. Podría ser que no haya una propiedad física F tal que podamos decir de una instancia c de la propiedad de tener cierto contenido mental, que c es idéntica a una instancia de F .

La tesis de fondo que se defenderá este trabajo es justamente, la tesis fodoriana de que los estados psicológicos deben ser individuados en base a sus contenidos mentales, que la explicación psicológica de la conducta es esencialmente intencional. La explicación psicológica es constitutivamente la explicación de la conducta (entendida en un sentido amplio, no sólo referida a la conducta externamente observable) a partir de estados de actitud proposicional como creencias y deseos que son individuados en base a componentes que, como los conceptos, son a su vez

individuados por sus contenidos. Llamaré a esta tesis, la “tesis representacionalista sobre los estados psicológicos”. El propósito central de este trabajo es profundizar la tesis representacionalista elucidando la noción de tener un contenido mental, de manera tal que dicha tesis sea aceptable desde el punto de vista del estándar fisicalista.

Si bien el proyecto de caracterizar a los estados mentales en términos de una noción semántica de información tiene su origen en el trabajo de Dretske (1981), considero que fue Fodor (por lo menos desde Fodor 1994) quien la empleó explícitamente para el propósito mencionado de defender una psicología intencional desde la perspectiva fisicalista. Fodor ubicó explícitamente al enfoque informacional como una pieza metafísica clave para la naturalización de la mente, para ofrecer una noción de contenido compatible con el criterio fisicalista. El aporte central de Dretske es adaptar la noción matemática información propuesta por Shannon y Weaver (1949) a una teoría de los estados mentales. En particular, Dretske ofrece una noción de *contenido informacional*, de qué es para una estructura s portar la información específica de que p , la información de que un hecho p ocurre. A partir de esta noción desarrolla una elucidación de las nociones que refieren a diferentes tipos de estados mentales, como la noción de representación perceptual, la noción de concepto, de creencia y de conocimiento. Ha habido un extenso debate en torno a la noción semántica de información, se han presentado una serie de problemas para la posibilidad de una noción de este tipo (para un abordaje sistemático de los problemas más ampliamente discutidos, ver Prinz 2002, capítulo 9 y Adams y Aizawa 2010). Sin embargo, considero que es Fodor quien ofreció un diagnóstico más claro del enfoque informacional en relación con la tarea específica de coordinar una psicología intencional con el fisicalismo.

Podemos distinguir dos problemas asociados a esta tarea. En primer lugar, tenemos que mostrar que la noción de contenido mental puede ser adecuadamente elucidada en términos de una noción semántica de información. En segundo lugar, tenemos que mostrar que la noción de contenido mental así elucidada es al menos compatible con el fisicalismo respecto de los estados psicológicos. Si bien, hasta donde se, Fodor no ha distinguido explícitamente estos dos aspectos de la tarea de naturalización, sí los ha tratado extensamente de forma separada.

Según Fodor (1998) las teorías informacionales sostienen que la relación entre los pensamientos y el mundo que determina el contenido de los primeros es una relación nomológica. Dicho de modo general, una representación R tiene a una propiedad P como contenido sólo si hay una ley que

determina que las instanciaciones de P causan instanciaciones de R (en ciertas condiciones C a especificar) (Fodor 1998, p. 12). Fodor (por ejemplo, 1994, 1995, 2008) considera que una teoría informacional del contenido mental tiene dos ventajas relacionadas con los dos problemas mencionados respecto de la naturalización de la mente (i. e., mostrar que el contenido puede ser caracterizado en términos de información y que esta caracterización es compatible con el fisicalismo). En primer lugar, considera que es la única teoría que puede ofrecer una caracterización adecuada de la noción de contenido mental. Esto es así justamente porque lo que considera la única alternativa naturalista relevante, alguna variedad de funcionalismo, no permite caracterizar adecuadamente la noción de contenido. Si bien nos adentraremos en este problema con más detalle en el capítulo siguiente, el principal problema que Fodor atribuye al funcionalismo es el holismo. El funcionalismo debe afirmar, según Fodor, que el contenido de una representación es determinado por *todas* las relaciones inferenciales o causales que mantiene otras representaciones. Esto conlleva al menos dos problemas graves sobre los que profundizaremos más adelante: La imposibilidad de cumplir con los requisitos de publicidad y composicionalidad para los conceptos, los componentes de los estados de actitud proposicional que necesita una psicología intencional. Por el contrario, como según el enfoque informacional, el contenido de una representación es determinado sólo por la relación que una representación tiene con una propiedad externa y no por ninguna relación que tenga con otra representación, Fodor considera que el enfoque informacional es atomista (el contenido de una representación es determinado con independencia del contenido de otras) y este rasgo permite evitar el problema del holismo. De este modo, suponiendo que los estados psicológicos son intencionales, son individuados en virtud de los contenidos de sus componentes, los conceptos, la teoría informacional parece ser la alternativa más adecuada para responder a la pregunta de qué es el contenido mental.

A su vez, como el contenido es determinado por una relación causal o nomológica entre un particular y diferentes instancias de un tipo, parece ser que el enfoque informacional ofrece condiciones de individuación para los contenidos mentales compatibles con una perspectiva fisicalista. Las relaciones causales nomológicas son propiedades que podemos atribuir a tipos físicos. Es por estos motivos que el enfoque informacional parece ofrecer una teoría del contenido mental que, por un lado, *se adecua a nuestra noción intuitiva de contenido* y, por otro lado, *caracteriza al contenido de un modo compatible con el fisicalismo*. Es decir, permite defender la idea de que los estados mentales, caracterizados según la tesis representacionista, son reales desde el punto de vista fisicalista.

Sin embargo, defender que la teoría informacional puede responder adecuadamente a estos dos

problemas (caracterizar la noción de contenido y mostrar que esta caracterización es fisicalista) involucra ciertos desafíos. Fodor mismo ha señalado originalmente lo que considero que son dificultades medulares que debe enfrentar esta defensa. En primer lugar, el enfoque informacional, al ser referencialista, al identificar el contenido con la referencia, no permite caracterizar a los estados como teniendo lo que comúnmente se ha llamado “contenido cognitivo”, una dimensión del contenido que se considera que los estados mentales deben poseer. Esto parece poner en duda la tesis de que la información puede caracterizar al contenido mental (por ejemplo, Fodor 2008). En segundo lugar, es ampliamente aceptado que los procesos cognitivos son procesos que se realizan sobre una base computacional, que la base física que permite naturalizar a los estados cognitivos es computacional. Si las instancias de procesos cognitivos son idénticas con instancias de procesos físicos, entonces lo son con instancias de procesos computacionales. Sin embargo, como los estados informacionales son definidos por propiedades extrínsecas y los estados computacionales son definidos por medio de propiedades intrínsecas, Fodor considera que la relación entre estados informacionales y computacionales es metafísicamente contingente. Es decir, la instanciación de un estado informacional no implica la de un estado computacional ni viceversa. Como señala Fodor (1994), esto es incompatible con la noción estándar de realización. Esta requiere una relación metafísicamente más fuerte entre el realizador y lo realizado. Por lo tanto, si caracterizamos al contenido en términos informacionales, parece que tampoco podemos dar cuenta de cómo se vincula con lo físico, o en qué sentido sus instancias podrían ser idénticas a instancias de una propiedad física.

Fodor ha ofrecido diversos argumentos para resolver estos dos problemas centrales del enfoque informacional. En primer lugar, dado que la teoría informacional del contenido no puede dar cuenta del contenido cognitivo, propone que una dimensión no semántica de las representaciones mentales, la forma sintáctica, cumpla la misma función que este tipo de contenido. De este modo la teoría informacional no dejaría sin explicar ningún aspecto relevante de nuestra noción intuitiva de representación mental. En segundo lugar, en relación el problema de la realización de los estados informacionales, Fodor ha argumentado que la noción estándar de implementación o realización es demasiado demandante. Los relata de este tipo de relación pueden guardar una relación metafísicamente contingente siempre que tengan algún tipo de relación nomológica (Fodor 1995). Que la instanciación del tipo realizado implique la instanciación del tipo realizador en los mundos nomológicamente posibles.

Si bien, como veremos a lo largo de los capítulos que siguen, considero que los aportes de

Fodor respecto de los problemas que involucran las tesis de la caracterización informacional del contenido y su realización son muy valiosos y desempeñan un papel central en lo que considero que es una solución adecuada para los mismos, creo que queda aún mucho trabajo por hacer para ofrecer una defensa acabada de dichas tesis.

2. Breve mapa del enfoque informacional

La tesis está organizada en dos grandes partes. Cada una de ellas se ocupa de uno de los dos problemas mencionados en la sección anterior: El de ofrecer una caracterización informacional del contenido mental y el de mostrar que el contenido así caracterizado se realiza sobre una base computacional y neuronal.

La primera parte se ocupa de proponer una teoría informacional de tres tipos de estados mentales relevantes. En primer lugar (capítulos 2 y 3), ofrezco una teoría del contenido intencional y cognitivo de los conceptos. En segundo lugar, determino la relación entre la noción de representación perceptual (que considero que debe ser definida independientemente de una semántica específica) y la de información semántica (capítulo 4). Por último (capítulo 6) ofrezco una teoría informacional del conocimiento, debatiendo con la caracterización clásica de Dretske (1981) sobre la base de discusiones recientes en teoría del conocimiento.

Además de determinar vínculos entre estos tipos de estados mentales y la noción de información, muestro cómo esta caracterización determina vínculos entre dichos estados. En primer lugar, defenderé que ser una representación conceptual implica no ser una representación perceptual. Es decir, defenderé una forma radical de *anti-empirismo de conceptos* (capítulo 5). No sólo digo que algunos conceptos no son representaciones perceptuales, sino que ninguno lo es. En segundo lugar defenderé que la noción de conocimiento implica que no hay conocimiento perceptual. Esta es una forma radical de *anti-empirismo*. En este caso se trata de un *anti-epistémico*. No se trata de que no todo conocimiento sea perceptual, sino de que *ningún caso de conocimiento es un caso de conocimiento perceptual*.

La segunda parte se ocupa de determinar relaciones conceptuales y metafísicas entre los contenidos mentales (entendidos informacionalmente) los tipos de estados físicos en los que se supone

que se realizan: los estados computacionales y neuronales. En primer lugar, elaboro diferentes consideraciones en torno a elucidaciones recientes de la noción de computación y su relación con la noción de procesamiento neuronal (capítulo 7). Presento y adopto la noción de computación de Piccinini y Bahar (2013). Muestro que la distinción entre esta noción y la noción de información con la que comúnmente es confundida es más fuerte de lo que estos autores creen. No sólo se da el caso de que computación no implica información sino que información tampoco implica computación. Además muestro de esta caracterización de la computación se sigue que el vínculo entre la computación y el procesamiento neuronal es más estrecho que lo que comúnmente se cree. El procesamiento neuronal no es una base de realización para un proceso computacional sino que es un tipo específico de proceso computacional, frente a la computación digital y la analógica.

En segundo lugar, sobre la base de la noción de computación propuesta, determino la relación que hay entre esta noción y la noción de proceso cognitivo, informacionalmente entendida (capítulo 8). Considero que a pesar de que la noción fodoriana de computación debe ser abandonada a favor de la de Piccinini y Bahar puede ofrecerse una defensa de la tesis fodoriana de que los estados cognitivos y los computacionales tienen una relación metafísicamente contingente. Considero además que si bien la noción que Fodor emplea para caracterizar a la computación, la noción de forma sintáctica, no puede ser empleada para este propósito, sí puede formar parte de la caracterización informacional de la cognición.

En tercer lugar, muestro que la relación que hay entre la noción de procesamiento neuronal, entendida a partir de la función misma del sistema nervioso, guarda una relación más estrecha con la noción de procesamiento cognitivo que la de computación (capítulo 9). Considero que si caracterizamos al procesamiento neuronal no sólo en base al tipo de vehículo que manipula sino también en base a la función del sistema que lo lleva a cabo, se sigue que hay una relación de implicación de procesamiento neuronal a procesamiento de información. Considero que la función específica del sistema nervioso, el control por retroalimentación negativa, implica a la noción de cognición. Antes de argumentar a favor de esta relación de implicación, me introduciré en un amplio debate desarrollado en torno a la noción de retroalimentación negativa y defenderé un modo peculiar de entenderla

En el último capítulo considero la noción de realización (capítulo 10). Comúnmente se considera que la relación que hay entre el procesamiento cognitivo y su base computacional y neuronal

es una relación de realización. Sobre la base de los resultados obtenidos en los capítulos anteriores y la noción de realización defendida en este capítulo se determinará si es posible afirmar que se dan las pretendidas relaciones de realización entre cognición, computación y procesamiento neuronal, y de qué tipo de realización se trata. Consideraré los dos principales tipos de enfoque sobre la realización, el enfoque dimensionado y el enfoque plano y argumentaré que si bien (como argumentan diversos autores) debemos abandonar el enfoque dimensionado, los argumentos que motivan dicho abandono implican también el abandono de aspectos constitutivos del enfoque plano.

La noción de realización resultante nos muestra que si bien la cognición se realiza en el mismo sentido sobre una base computacional y sobre una base neuronal se realiza también de diferente modo sobre ellas. Aunque la instanciación de un proceso neuronal sea metafísicamente suficiente para la instanciación de un proceso cognitivo y la instanciación de un proceso de computación neuronal sea sólo nomológicamente suficiente para la instanciación de un proceso cognitivo, la cognición se realiza en el mismo sentido en ambos tipos de proceso. Además, confirmaré ahora sobre la base de la elucidación de la realización, la tesis defendida en el capítulo 7 de que lo computacional no se realiza sobre lo neuronal.

Este capítulo nos muestra, finalmente, que la noción de información no sólo nos permite caracterizar adecuadamente a los diferentes tipos de representaciones mentales sino que también nos permite explicar en qué sentido los estados intencionales mentales se realizan sobre una base física, computacional y neuronal y por lo tanto son aceptables, desde el punto de vista del criterio fisicalista o naturalista.

Considero que la noción de información puede funcionar como la idea, tomada en un sentido bastante metafórico respecto de la cita de Kant, que revele la trama de relaciones entre los diferentes componentes que constituyen la mente. La noción de información nos permite caracterizar a los diferentes tipos de estados mentales y defender tesis sustantivas sobre las relaciones que guardan entre sí y sobre las diferentes relaciones que guardan con su base de realización.

Parte I:

La teoría informacional del contenido mental

“En el principio era la información. El Verbo vino luego. La transición fue lograda por el desarrollo de organismos con la capacidad de explotar selectivamente esta información para sobrevivir y perpetuar a su especie” Dretske (1981)

Como adelanté en el capítulo anterior, en esta primera parte me ocuparé de la caracterización informacional de determinados tipos de estados mentales que individuamos en virtud de sus contenidos, estados que entrarían en una psicología intencional. Como vimos, esta es la primera parte de una defensa de un enfoque informacional de los estados mentales. Aquí mostraré que diferentes estados mentales intencionales pueden efectivamente caracterizarse en términos informacionales. En la segunda parte mostraré que los estados mentales así caracterizados nos permiten ser fisicalistas respecto de los mismos, es decir, hace que los estados intencionales sean científicamente aceptables. En esta primera parte no me ocuparé de todos los estados intencionales sino sólo de tres tipos de ellos: Los conceptos, las representaciones perceptuales y el conocimiento.

En los capítulos 2 y 3 me ocuparé de defender una peculiar concepción informacional de los conceptos. Esta concepción tendrá dos particularidades centrales. En primer lugar, ofreceré una caracterización informacional no sólo del contenido intencional sino también del llamado “contenido cognitivo”. A partir de Fodor (por ejemplo 2008) se cree que la semántica informacional no puede dar cuenta del contenido cognitivo y que la única alternativa relevante para dar cuenta del mismo es la semántica de rol funcional, que Fodor considera que adolece problemas insuperables. La segunda peculiaridad de mi propuesta es que integra componentes informacionales y funcionales. Esta integración hace posible que el enfoque informacional explique el contenido cognitivo y que los roles funcionales eviten ciertos problemas que se les han atribuido comúnmente.

En el capítulo 4 ofreceré una concepción de las representaciones perceptuales que no apela a ningún tipo de propiedad semántica o sintáctica sino sólo a la facultad a la que pertenecen. Para ofrecer tal caracterización llevo adelante una tarea de elucidación de la noción de sistema perceptual. Curiosamente, la caracterización de los sistemas perceptuales ofrecida implica que las representaciones que pertenecen a estos sistemas poseerán un contenido informacional.

En el capítulo 5, a partir de haber atribuido un contenido informacional tanto a las representaciones perceptuales como a las conceptuales, considero el tipo de relación que guardan estos dos tipos de representaciones. Argumentaré a favor de cuatro tesis respecto de la relación entre lo

conceptual y lo perceptual. Diré que ser conceptual no implica ser perceptual, que ser perceptual no implica ser conceptual, que ser no perceptual no implica ser conceptual y que ser conceptual implica ser no conceptual. Esta última tesis constituye un anti-empirismo respecto de los conceptos. En el capítulo 3 argumentaré en contra de la concepción informacional que Prinz propone sobre los conceptos. Sin embargo, la nota más saliente del enfoque de Prinz es que considera que todos los conceptos son representaciones perceptuales. Por el contrario, mi propuesta informacional implica no sólo la negación de esta tesis, que no todo concepto es una representación perceptual, sino la tesis más radical de que ser una representación conceptual implica no ser una representación perceptual.

Por último, en el capítulo 6, me ocuparé de proponer una caracterización informacional del conocimiento y defender una tesis anti-empirista respecto de la misma noción. Para llegar a una teoría informacional del conocimiento me introduciré en debates contemporáneos en teoría del conocimiento. En particular, defenderé un enfoque de la seguridad, en contra de las llamadas “epistemologías de la virtud” y en contra del confiabilismo de procesos. Considero que podemos formular una caracterización informacional del conocimiento equivalente a la condición del enfoque de la seguridad. Por el contrario, considero que la caracterización informacional clásica de Dretske adolece de problemas semejantes a los que padece la epistemología de la virtud. Por último, mostraré que la noción informacional de conocimiento propuesta implica la tesis de que no hay conocimiento perceptual. Implica un anti-empirismo respecto del conocimiento.

Capítulo 2:

Una teoría informacional para los casos Frege

1. Introducción

Mi objetivo en este capítulo es ofrecer una propuesta informacional del contenido cognitivo. Sin embargo, mi defensa se aplica de modo general a cualquier teoría causal. Las teorías causales del contenido mental son una familia de teorías muy diversas cuyo denominador común es la idea de que el contenido de una representación mental es determinado por algún tipo de relación causal que guarda con un objeto o propiedad distal (Adams y Aizawa 2010). La teoría informacional es una propuesta de tipo causal. Sin embargo, como el problema que consideraré en este capítulo es un problema general de toda propuesta causal, intentaré ofrecer una solución a dicho problema que valga para cualquier enfoque causal y no sólo para las teorías informacionales.

La expresión “teoría causal” es equívoca ya que, como veremos, la semántica de rol funcional en su variante naturalista también sostiene que las representaciones adquieren su contenido por medio de relaciones causales. La diferencia es que en el caso del funcionalismo se trata centralmente de relaciones causales entre representaciones. Sin embargo, como esta es la terminología estándar (Adams y Aizawa 2010) y Fodor (2008), quien protagonizará la discusión de este capítulo, la emplea, me ceñiré a ella. Fodor (2008) se compromete con una teoría causal de la referencia de las representaciones mentales. Aquí intentaré usar los recursos de este tipo de propuesta para explicar el contenido cognitivo. Para entender la relevancia de un enfoque de este tipo debo comenzar describiendo la discusión de la que emerge.

Los conocidos casos ofrecidos por Frege (1892) para argumentar a favor de su caracterización

de la noción de significado han sido empleados generalmente para argumentar sobre la naturaleza del contenido mental y, más específicamente, sobre el contenido de los conceptos mentales. Se ha considerado usualmente que este tipo de casos motivan la introducción de una dimensión del contenido que va más allá de la referencia. Esto es lo que Frege ha llamado “sentido” y lo que en filosofía de la mente se ha llamado frecuentemente “contenido cognitivo” o “contenido estrecho”. Jerry Fodor ha ofrecido en diversos trabajos un diagnóstico peculiar sobre este tipo de casos. Fodor considera que el modo típico en que se han explicado es apelando (explícita o implícitamente) a una teoría de rol funcional. Sin embargo, también ha argumentado extensamente que estas teorías adolecen de un grave y aparentemente inevitable problema: el holismo semántico. Por otro lado, Fodor considera que una propuesta alternativa a la de rol funcional, algún tipo de teoría causal, tiene un atractivo importante en la medida en que ofrece una caracterización naturalista del contenido mental, pero principalmente porque, al ser atomistas, no adolecen el problema del holismo semántico. Sin embargo, este tipo de teoría, al identificar el significado con la referencia, no cuenta con los recursos para explicar los casos Frege. Esto parece dejarnos en medio de un dilema: o bien explicamos los casos Frege y caemos en el holismo o bien evitamos el holismo pero no explicamos los casos Frege. La original respuesta de Fodor a este dilema, a través de diversos trabajos, es considerar que los casos Frege no necesitan tener una explicación semántica. Podemos explicarlos en base a propiedades puramente sintácticas de las representaciones mentales. Si bien Fodor ha seguido esta línea de argumentación a lo largo de diferentes trabajos sin variaciones relevantes para la discusión que quiero abordar aquí (por ej. Fodor 1994 y 1998), en la exposición de su propuesta me ceñiré principalmente a Fodor (2008), ya que el autor afirma allí explícitamente que no adhiere a sus soluciones previas al problema de Frege.

En este capítulo intentaré ofrecer una salida alternativa al dilema fodoriano. Propondré una explicación semántica de los casos Frege que incluye elementos de las dos teorías involucradas en el dilema. Considero que los roles funcionales permiten construir una teoría causal que no sea referencialista y al mismo tiempo la teoría causal permite evitar el holismo de los roles funcionales.

El capítulo tendrá la siguiente estructura. En primer lugar, en la sección 2, ofreceré una caracterización más detallada de lo que antes mencioné como “dilema de Fodor” como marco del debate y de la solución sintáctica fodoriana. Luego, en la sección 3, intentaré bloquear una de las consecuencias problemáticas del holismo de los roles funcionales. Argumentaré que aún si consideramos que los roles funcionales son holísticamente determinados, pueden combinarse composicionalmente. El holismo sin embargo, es también problemático porque impide que el

funcionalismo satisfaga el requisito de publicidad para los conceptos. Considero que este problema puede resolverse si combinamos a los roles funcionales con ciertos elementos de una semántica de tipo causal. En la sección 4 describiré con algo de detalle la propuesta resultante de integrar estos elementos y el modo en que se resuelve el problema de la publicidad.

2. El dilema de Fodor y la estrategia sintáctica

Como he adelantado en la introducción, llamo “dilema de Fodor” a un problema que enfrentan las teorías del contenido mental para explicar los casos de Frege. Básicamente la idea es que la única alternativa relevante para explicar los casos Frege son las semánticas de rol funcional. Pero estas adolecen del problema del holismo semántico y por lo tanto son inaceptables. Por otro lado, las teorías causales, al ser atomistas, parecen ser las propuestas adecuadas para evitar el problema del holismo, pero como son referencialistas (identifican el contenido con la referencia) no cuentan con los recursos para explicar los casos Frege. Por lo tanto, si explicamos los casos Frege, caemos en el holismo. Pero si evitamos el holismo, no seremos capaces de explicar los casos Frege. En lo que sigue profundizaré sobre cada uno de los elementos de este dilema y presentaré brevemente la solución fodoriana. Para ello me basaré exclusivamente en Fodor (2008) ya que es la obra más reciente en la que encontramos el problema y la solución planteados exhaustivamente y Fodor reniega explícitamente de sus soluciones previas (Fodor 2008, p. 57).

Los casos Frege son aquellos casos en los que la substitución de expresiones de un lenguaje que tienen la misma referencia no preserva el valor de verdad de la fórmula de la que forman parte. Esto sucede principalmente en los llamados contextos lingüísticos opacos de substitución. Estos incluyen, en particular, los contextos de atribución de actitudes proposicionales (como creencias o deseos). Por ejemplo, aunque sea verdadera la oración “Juan cree que la estrella vespertina es bella”, puede no ser verdadera la oración “Juan cree que la estrella matutina es bella”, aún cuando “estrella vespertina” y “estrella matutina” tengan la misma referencia. Aunque Frege haya planteado este problema en el marco de la caracterización de una noción no psicológica del contenido, se ha considerado frecuentemente que los casos Frege introducen restricciones para construir una semántica de las representaciones mentales. La explicación psicológica necesita que la semántica de la representación mental dé cuenta del hecho de que Juan puede creer que la estrella matutina es bella, puede tener esta

actitud proposicional, sin creer que la estrella vespertina es bella (Fodor 2008, p. 52). Se ha considerado usualmente que lo que los casos Frege muestran es que *el contenido* de las representaciones mentales no puede ser identificado con la referencia. Si alguien puede tener la representación ESTRELLA MATUTINA sin tener la representación ESTRELLA VESPERTINA, entonces, dado que estas representaciones tienen la misma referencia, el contenido de las mismas no puede identificarse con aquella referencia.

Según Fodor (2008), la alternativa más común a una teoría referencialista del significado (una que identifique el contenido con la referencia), es una teoría de dos factores donde el contenido sea caracterizado por medio de la referencia y otra dimensión semántica que puede llamarse “sentido”. Fodor considera que hoy en día los sentidos son ampliamente identificados con roles funcionales y por lo tanto la teoría que puede dar cuenta de ellos es la llamada “semántica de rol funcional” (Fodor 2008, p. 53). Dicho muy brevemente, según esta propuesta, el contenido de una representación está determinado por las relaciones causales o inferenciales que tiene con otras representaciones. Ahora bien, Fodor ha argumentado extensamente que este tipo de propuestas adolece del problema del holismo (por ej., Fodor y Lepore 1992 y Fodor y Lepore 2002). Básicamente, el problema es que el criterio que necesitamos para determinar cuáles de las inferencias o relaciones causales en las que entra una representación determinan su contenido involucra la distinción analítico/sintético criticada por Quine (1953). Como no podemos admitir este criterio, debemos identificar al contenido de un concepto con su rol funcional total, con *todas* las inferencias en las que figura. Esto constituye un problema principalmente por dos razones. En primer lugar, si la semántica es holista en el sentido descripto, entonces dos personas que difieran en alguna de esas inferencias (o peor aún, una misma persona en diferentes momentos) no tendrán el mismo concepto. Pero es plausible que estas diferencias se den masivamente. Esto quiere decir que el requisito de publicidad para una teoría de conceptos, la exigencia de que los conceptos sean compartidos, no puede ser satisfecho por esta propuesta. Más aún, esto implica que la teoría de rol funcional es incompatible con generalizaciones psicológicas, con la formulación de las leyes intencionales que constituyen la explicación psicológica de la conducta, según Fodor (1994). Por otro lado, Fodor también ha argumentado que una semántica holista tampoco es apta para satisfacer el requisito de composicionalidad para una teoría de conceptos. Si los conceptos son identificados con roles funcionales totales, entonces los conceptos no se combinan composicionalmente, es decir, el significado de un concepto complejo no puede reducirse al significado de sus partes. El ejemplo clásico de Fodor es el caso en el que el concepto PELIGROSA forma parte

del rol funcional del concepto complejo VACA MARRÓN sin que PELIGROSA forme parte del rol funcional de VACA o de MARRÓN (Fodor y Lepore 1992 y Fodor y Lepore 2002).

Frente a este problema, la propuesta alternativa que Fodor ha favorecido en diversas ocasiones es una semántica de tipo informacional. Para una primera y esquemática aproximación, las propuestas informacionales sostienen que una representación R representa a una propiedad P si y sólo si hay una ley que determina que los casos de P causen casos de R (Fodor 1998). Fodor (2008) no adhiere específicamente a una teoría informacional pero sí se compromete con la familia de teorías de la que esta teoría forma parte. Afirma que algún tipo de teoría causal es verdadera (Fodor 2008, p. 199) e incluso se toma un capítulo entero (capítulo 7) para responder a las objeciones tradicionales que se le han presentado a este tipo de enfoque. Estas propuestas tienen, según Fodor, dos ventajas cruciales. En primer lugar, parece ser una de las pocas teorías aptas para naturalizar el contenido mental. A pesar de que este es un aspecto muy relevante de las propuestas causales, no me detendré en él ya que no es indispensable para el problema que me ocupa. La segunda ventaja que Fodor atribuye a las teorías causales, y la que resulta relevante para el problema planteado, es su atomismo. Vimos que según esta propuesta, el contenido de una representación R de un sujeto S es determinado puramente por la relación causal que tiene con una propiedad distal. Esto quiere decir que las otras representaciones que S posea y las relaciones (inferenciales o causales) que tengan con R no afectan de ningún modo el contenido de R . En esto consiste el atomismo de la propuesta causal y es esta característica la hace inmune al holismo semántico. No tenemos que determinar qué inferencias son relevantes para determinar el contenido de una representación porque ninguna lo es.

Ahora bien, a pesar de estas ventajas, hay un aparente punto débil de las teorías causales. Fodor sostiene que estas propuestas son referencialistas. Esto quiere decir que, según este enfoque, las expresiones que tengan la misma referencia serán sinónimas, tendrán el mismo significado. Por ejemplo, si adoptamos una semántica informacional y si asumimos que h_2o y agua son la misma propiedad, entonces si el concepto AGUA está en una relación causal nomológica con agua, también lo estará con h_2o y si el concepto H_2O está en una relación causal nomológica con h_2o también lo estará con agua. Por lo tanto, los conceptos AGUA y H_2O serán sinónimos (Fodor 1998). Ahora bien, como vimos, se considera comúnmente que los casos de tipo Frege son contraejemplos para una semántica referencialista. El contenido no puede identificarse con la referencia.

Si Fodor está en lo correcto con lo afirmado hasta ahora (en este trabajo asumiré que sí) la idea

de que los casos Frege requieren una explicación semántica nos hace enfrentarnos a un dilema. Por un lado, si adherimos a alguna teoría causal del contenido evitamos el holismo pero no podemos explicar los casos Frege. Por otro lado, si adherimos a una semántica de rol funcional podemos explicar los casos Frege pero no podemos evitar el holismo semántico.

La respuesta fodoriana a este dilema, a través de diferentes trabajos, (por ej., 1994, 1998, 2008) es la idea de que los casos de Frege no precisan la introducción de una nueva dimensión del contenido conceptual sino que pueden ser explicados a partir de propiedades no semánticas, puramente sintácticas de los conceptos. Fodor (2008) acepta que los casos Frege muestran que los conceptos no pueden ser individuados sólo en base a la referencia, pero afirma de ello no se sigue que aquello adicional que precisamos para la individuación sea un parámetro del contenido. Fodor (2008) divide su explicación de los casos Frege en dos partes. En la primera considera el caso de los conceptos complejos y en la segunda se ocupa de los conceptos básicos.

Los conceptos complejos son aquellos que poseen una estructura constituyente. Los constituyentes de una representación son sus partes semánticamente interpretables. La idea de Fodor es que dos conceptos complejos que tienen la misma referencia pero diferentes constituyentes tendrán diferentes condiciones de posesión. Por ejemplo, uno puede tener el concepto complejo ESTRELLA MATUTINA pero no ESTRELLA VESPERTINA si uno carece del concepto VESPERTINA. Pues poseer VESPERTINA es una de las condiciones para poseer el concepto complejo ESTRELLA VESPERTINA pero no para poseer ESTRELLA MATUTINA. De este modo explicamos que alguien pueda poseer uno de dos conceptos que tienen la misma referencia sin poseer el otro. Esto es una explicación sintáctica de los casos Frege para los conceptos complejos porque los constituyentes, si bien son partes interpretables semánticamente, son constituyentes sintácticos (es decir, que pueden ser individuados sintácticamente, sin apelar a propiedades semánticas) (2008, p. 61).

Antes de considerar el caso de los conceptos simples, es importante aclarar algo sobre los constituyentes los conceptos complejos que será importante más adelante. Por supuesto, cuando Fodor habla de constituyentes de un concepto complejo, está hablando de partes contenidas en el concepto, de su *estructura interna*. En particular, un concepto *R* sólo relacionado inferencialmente con otro *C* no es un constituyente de *C*. Por ejemplo, PERRO MARRÓN no es un concepto que tenga como constituyente a ANIMAL, aunque esté relacionado inferencialmente con él. Pero sí tiene como constituyente a PERRO (con el cual está además inferencialmente relacionado), ya que PERRO forma

parte de la estructura interna de PERRO MARRÓN. Esto es importante para notar que la admisión de conceptos complejos no implica la negación del atomismo respecto de los conceptos. Si bien necesito tener PERRO y MARRÓN para tener el concepto PERRO MARRÓN, no necesito ningún otro concepto para tener PERRO, o cualquier otro concepto básico (o léxico), es decir, un concepto sin diferentes componentes sintácticos. El atomismo es una tesis sobre los conceptos básicos. Este punto será relevante cuando, en la sección siguiente, argumente a favor de la composicionalidad de los roles funcionales.

Ahora, para el caso de los conceptos básicos, aquellos que no tienen una estructura constituyente, también pueden darse los casos de Frege. Fodor considera que los mejores candidatos a ser conceptos básicos son los expresados por nombres monomorfémicos de individuos. Por ejemplo, CICERÓN y TULIO pueden ser conceptos básicos que si bien tienen la misma referencia parece posible tener uno sin tener el otro. Si bien Fodor considera otros casos, no se compromete con otros tipos de conceptos simples.

Como hemos visto, Fodor considera que lo que los casos Frege motivan la introducción de un criterio de individuación de conceptos que permita que conceptos con igual contenido sean diferentes conceptos. Necesitamos decir que son diferentes conceptos, según Fodor, porque pueden tener diferentes poderes causales. Por ejemplo, creer que Cicerón era gordo puede causar que afirme “Cicerón era gordo” pero creer que Tulio era gordo no. La propuesta de Fodor es emplear la tesis de que los procesos mentales son computacionales (lo que llama “teoría computacional de la mente”) para proporcionar un criterio de individuación de conceptos que explique estos casos. La teoría computacional afirma, según Fodor, que el aspecto causalmente relevante de un concepto (o de la fórmula del lenguaje del pensamiento de la que forma parte) no es su contenido sino su sintaxis. Si esto es así, se explica que conceptos con el mismo contenido puedan tener poderes causales diferentes. Simplemente tendrán propiedades sintácticas diferentes. Pero esta diferencia sintáctica no es una diferencia en constituyentes, ya que se trata de conceptos básicos, sino de *forma*. La forma sintáctica de una representación es simplemente alguna propiedad física (a Fodor no le interesa determinar específicamente qué es) a la que es sensible la computación mental (o neuronal).

Hasta aquí hemos considerado el problema que plantea Fodor respecto de la explicación de los casos Frege por parte de una teoría de conceptos y la solución ofrecida. En lo que sigue intentaré proponer otra solución para el dilema de Frege. Tomaré la opción de defender una explicación

semántica empleando roles funcionales, pero intentaré resolver el problema del holismo apelando a las herramientas de la otra teoría que figura en el dilema: la teoría causal.

3. Composicionalidad

Hemos visto que, según Fodor, el problema principal de las propuesta funcionales es su holismo. Ahora bien, el holismo es un problema porque tiene dos consecuencias indeseables. Si los roles funcionales se determinan holísticamente entonces no se combinan composicionalmente y no son ampliamente compartidos. En esta sección me ocuparé del problema de la composicionalidad. Intentaré mostrar que aún si los roles funcionales implican el holismo, se combinan composicionalmente.

Fodor (2008) no desarrolla el problema que los roles funcionales tienen con el requisito de composicionalidad sino que simplemente nos remite a Fodor y Lepore (2002). Fodor y Lepore (2002) afirman que la composicionalidad es una propiedad que un sistema de representaciones tiene cuando (i) contiene símbolos primitivos y símbolos que son sintáctica y semánticamente complejos, y (ii) los últimos heredan sus propiedades semánticas y sintácticas de los primeros (2002, p.1). Los principales argumentos a favor de la composicionalidad de las representaciones mentales son el fenómeno de la productividad y la sistematicidad de los pensamientos. Dicho brevemente, la productividad es la propiedad que tiene un sistema de representaciones cuando puede formar infinitos símbolos semántica y sintácticamente diferentes a partir de un conjunto finito de primitivos. La sistematicidad es la propiedad de un sistema de representaciones de tener expresiones diferentes pero relacionadas semántica y sintácticamente. El argumento central de Fodor en contra de la composicionalidad de los roles funcionales puede resumirse del siguiente modo. Según la semántica de rol funcional, el significado de una expresión está constituido al menos en parte por sus relaciones inferenciales (o causales) con otras expresiones. Pero, no hay un principio (como la distinción analítico/sintético) que permita determinar qué inferencias son constitutivas del significado de una expresión y cuáles no. Por lo tanto, el significado de una expresión debe ser determinado, según este enfoque, por *todas* las inferencias en las que figura (Fodor y Lepore 2002, p. 23). Hasta aquí tenemos un argumento que va de la teoría de rol funcional al carácter holista de esta propuesta. Para llegar al requisito de composicionalidad debemos decir algo más. La idea de Fodor es que si tenemos una expresión compleja como VACA MARRÓN, su rol funcional puede involucrar inferencias que no forman parte

del rol funcional de VACA ni de MARRÓN. Por ejemplo, si creemos que las vacas marrones son peligrosas, el rol funcional de VACA MARRÓN (pero no los roles de VACA o de MARRÓN) involucrará el concepto PELIGROSA. Como no tenemos un modo de restringir las inferencias que constituyen el significado de una expresión compleja como esta, como el funcionalismo tiene que ser holista, no tenemos modo de decir que PELIGROSA no formará parte del significado de VACA MARRÓN. Como PELIGROSA no forma parte del significado de VACA o de MARRÓN, la propuesta de rol funcional falla en cumplir el requisito de composicionalidad. Por lo tanto, no podemos identificar a los contenidos conceptuales con roles funcionales. Llamaré a los casos de este tipo, “Casos Fodor”. Si los casos Frege muestran que no podemos individuar a los conceptos sólo por su referencia, los casos Fodor mostrarían que (asumiendo, con Fodor (2008), que la teoría de rol funcional es la única alternativa relevante para dar cuenta del contenido estrecho) *el componente adicional que requieren los casos Frege no puede ser semántico*.

Ahora bien, considero que una propuesta de rol funcional puede satisfacer el requisito de composicionalidad. Pienso que los casos Fodor dependen de una caracterización de los conceptos complejos que una propuesta funcional no está obligada a aceptar. Como veremos, esta solución no implica evitar el holismo. Como el holismo es además problemático para el requisito de publicidad, habrá que decir algo más para que una propuesta de rol funcional sea aceptable. La solución que intentaré ofrecer en esta sección al problema de la composicionalidad se aplica generalmente a las propuestas funcionales. Pero la solución que, en la próxima sección, ofreceré al problema del holismo y de la publicidad se aplica sólo a la propuesta que quiero ofrecer en este capítulo.

Los casos Fodor presuponen que los conceptos complejos tienen una semántica caracterizada funcionalmente. Esto es así porque estos casos funcionan en el contexto de una reducción al absurdo del funcionalismo. Fodor pretende mostrar que si suponemos que los contenidos conceptuales son roles funcionales entonces los conceptos no se combinan composicionalmente. Pero como sabemos que los conceptos sí se componen de este modo, podemos concluir que los contenidos conceptuales no son roles funcionales.

Además, considero que los casos Fodor suponen que parte de las condiciones de individuación de un concepto complejo es su estructura compleja *sintáctica*. Los casos Fodor son casos donde hay un concepto complejo *C* cuyo rol funcional no se reduce a los roles funcionales de los componentes sintácticos de *C*. Veamos por qué. En el caso de Fodor, el concepto que representa a las vacas marrones

tiene un rol funcional que incluye al rol de VACA y al de MARRÓN pero que no se reduce a ellos, su rol funcional incluye además el rol de PELIGROSA. Esto es un problema solamente porque, independientemente de este contenido identificado funcionalmente y que incluye tanto a VACA , a MARRÓN, como a PELIGROSA, sabemos que el concepto está compuesto *sólo* por VACA y por MARRÓN. Pero para poder afirmar que estos son sus únicos componentes no podemos basarnos obviamente en el rol funcional del concepto (puesto que este involucra también a PELIGROSA) sino en algún otro aspecto suyo. Teniendo el trasfondo de la propuesta fodoriana, el único candidato relevante parece ser su estructura sintáctica. El concepto que refiere a las vacas marrones tiene una *estructura interna* que tiene como componentes *sintácticos* sólo a VACA y a MARRÓN y un rol funcional no se reduce al de estos dos conceptos.

Ahora bien, como vimos, los casos Fodor funcionan en el contexto de una reducción al absurdo de la teoría de rol funcional y por lo tanto deben presuponerlas. Pero más aún, *sólo* deben presuponerlas a ellas. Si hay supuestos ajenos a las teorías de rol funcional entonces el que defiende ese tipo de propuestas podría adjudicar el éxito del argumento a esos supuestos. En particular, una reducción al absurdo de la semántica de rol funcional no debe involucrar criterios de individuación de conceptos que sean ajenos a este tipo de propuestas. Pero como vimos, las teorías de rol funcional no están necesariamente comprometidas con la tesis de que la forma sintáctica de un concepto es parte de lo que debemos usar para individualarlo. El único aspecto de un concepto con el que este tipo de propuestas está estrictamente comprometido a emplear como criterio de individuación es el rol funcional. Por este motivo, considero que los casos Fodor sólo valen en contra del funcionalismo si pueden prescindir de dicho supuesto.

Pero si dejamos a un lado la forma sintáctica y nos atenemos a sólo al criterio de individuación por medio de roles funcionales, decir que un concepto está compuesto exclusivamente por VACA y por MARRÓN sólo puede querer decir que tiene exclusivamente los roles funcionales de estos conceptos. Si los conceptos son *identificados* con roles funcionales, no tiene sentido decir que un concepto está compuesto sólo por VACA y MARRÓN pero que su rol funcional es más que la suma de los roles de estos dos conceptos. Si no aceptamos otro aspecto para individualar a un concepto además del rol funcional, no tenemos modo de decir que el concepto del ejemplo de Fodor sólo está compuesto por VACA y por MARRÓN. Por lo tanto, dado que el supuesto fodoriano, al no ser un supuesto funcionalista, no puede figurar en una reducción del funcionalismo y dado que el argumento no es efectivo sin ese supuesto, entonces podemos concluir que los casos Fodor no muestran que los roles

funcionales no son composicionales.

Ahora bien, es cierto que si Fodor da motivos *independientes del argumento en contra de los roles funcionales* para afirmar que la forma sintáctica debe ser parte de las condiciones de individuación de un concepto, entonces el funcionalista debería aceptar dicha afirmación. Los casos Fodor no necesitan valerse sólo de supuestos funcionalistas. En tanto reducción al absurdo, el argumento que emplea estos casos debe valerse de supuestos funcionalistas, *más cualquier otra tesis motivada por razones independientes de esa reducción.*

Fodor ofrece razones para pensar que las representaciones tienen una forma sintáctica y que el pensamiento es un proceso que opera sobre ellas. Fodor da motivos para considerar que el pensamiento es un proceso computacional. Pero, como vimos en la segunda sección, la razón que ofrece Fodor para afirmar que la forma sintáctica no es meramente algo que las representaciones poseen sino algo que *constituye parte de sus condiciones de individuación* es precisamente el hecho de que esta es la única forma de explicar los casos Frege. *Pero la tesis de que esta es la única forma se sigue de que las teorías de rol funcional no son aceptables.* Como vimos, el rechazo de las semánticas funcionales depende, en parte, de los casos Fodor. Por lo tanto, sería circular que los casos Fodor presupongan que la forma sintáctica es parte de las condiciones de individuación de un concepto. Si la individuación en base a la forma sintáctica se sigue, en parte, de que la semántica funcional no cumple con composicionalidad, el argumento en contra de la composicionalidad no puede presuponer ese modo de individuación.

Resumiendo, he argumentado que los casos Fodor dependen de un supuesto. El supuesto es que los conceptos no son individuados sólo en base a los roles funcionales sino también por su forma sintáctica. Esto es así porque los casos Fodor surgen cuando el rol funcional del concepto complejo no se reduce a los roles de los componentes sintácticos del mismo. He argumentado que el supuesto no es un supuesto de los enfoques funcionalistas y por lo tanto no puede figurar en una reducción al absurdo del funcionalismo. Pero he mostrado que si prescindimos de este supuesto el argumento no es efectivo. Si individualizamos a los conceptos sólo en base a los roles funcionales, entonces estos son composicionales. Argumenté además que no sirve pensar que Fodor no necesita limitarse a supuestos funcionalistas, si ofrece razones generales para aceptar la forma sintáctica como condición de individuación. Pues parte de la razón que ofrece Fodor para aceptar esta condición es justamente que la semántica de rol funcional no cumple con composicionalidad. El argumento en contra de la

composicionalidad no puede presuponer dicha condición.

Por estos motivos considero que el argumento de Fodor en contra de la composicionalidad de los roles funcionales no es efectivo. Pasaré a considerar en la próxima sección el problema restante de las propuestas funcionales.

4. Holismo

Vimos que el argumento en contra de la composicionalidad de los roles funcionales no depende sólo de los casos Fodor sino también del argumento que va de la teoría de rol funcional al holismo. Considero que bloquear los casos Fodor basta para argumentar a favor de la composicionalidad de los roles funcionales. Pero como vimos, lo que argumenté en la sección anterior deja intacto al holismo de los roles. Esto es un problema porque el holismo no sólo colabora con el argumento en contra de la composicionalidad de los roles funcionales sino también con el argumento en contra de su publicidad.

Hemos visto que si los roles funcionales totales constituyen la identidad de un concepto entonces este no será compartido. Dos personas que difieran en alguna inferencia respecto de un concepto dado (o peor aún, una misma persona en diferentes momentos) no tendrán el mismo concepto. Pero es plausible que estas diferencias se den masivamente. Esto quiere decir que el holismo impide que una propuesta de rol funcional satisfaga el requisito de publicidad. Peor aún, esto implica que la teoría de rol funcional es incompatible con generalizaciones psicológicas, con la formulación de las leyes intencionales que constituyen la explicación psicológica de la conducta, según Fodor (1994). Si los conceptos no son compartidos entonces las explicaciones intencionales de nuestras conductas nunca pueden generalizarse. De este modo, debemos encontrar un modo de restringir la parte del rol funcional de un concepto que constituye su significado. Como vimos, dados los argumentos de Quine (1953), no podemos apelar a la distinción analítico/sintético. Pero este parecía el modo más natural de restringir el rol funcional.

Reelaborando una idea de Prinz (2002), propongo recurrir a la noción de *proxytipo por default* para restringir los roles funcionales de modo tal que puedan ser compartidos. Los proxytipos por default son representaciones que uno instanciaría si considerara una categoría sin un contexto determinado. Cuando un contexto no es especificado, usamos las representaciones *que figuran más*

frecuentemente en nuestras interacciones con los miembros de una categoría. Prinz considera que los proxytipos por default son relativamente estables, ampliamente compartidos y frecuentemente responsables de guiar nuestra conducta ordinaria respecto de los miembros de una categoría dada. Las representaciones incluidas en un proxytipo por default son las que más frecuentemente atribuimos a una categoría (Prinz 2002, cap. 11). Adaptando esta idea a una teoría de rol funcional podríamos decir que el contenido de un concepto está determinado por las inferencias que más frecuentemente hacemos en nuestras interacciones con los miembros de una categoría (y aquellas que haríamos si consideramos la categoría fuera de contexto). Llamaré a estas inferencias “inferencias por default” y diré que el contenido de un concepto está determinado por las inferencias por default en las que figura. Restringir los roles funcionales de este modo nos acerca un poco más a la posibilidad de compartirlos. Sin embargo, considero que esto no es aún suficiente. Veamos por qué.

Si consideramos con algo de detalle el modo en que se determina un rol funcional, veremos que la solución considerada no será útil para una propuesta funcionalista. Para determinar qué rol funcional desempeña un concepto dado, típicamente se emplea un proceso de ramsificación (Lewis 1970). Para ramsificar un concepto comenzamos listando las inferencias en las que ese concepto entra. Luego reemplazamos cada instancia de la palabra que se corresponde con ese concepto con una variable y usamos la descripción resultante para definir el concepto. Pero nuestra especificación del rol está aún incompleta porque no hemos definido los otros conceptos expresados por nuestra descripción de ese rol. *Todo otro concepto que aparezca en la descripción debe ser ramsificado a su vez.* Deben ser reemplazados por variables que serán definidas por las inferencias en las que aparecen. Por ejemplo, si CUATRO PATAS forma parte del rol funcional de PERRO entonces, también formarán parte del rol funcional de PERRO los conceptos que aparecen en las inferencias en las que figura CUATRO PATAS. Los conceptos de CABALLO, GATO, VACA, etc.

Es importante resaltar el hecho de que este último paso de la definición funcional es crucial para el funcionalismo. Pues si queremos decir que dos individuos tienen el mismo concepto *C* debemos especificar varias de las inferencias en las que aparece *C*. Estas inferencias se identifican en parte por medio de *los otros conceptos* que figuran en ellas (los conceptos que no son *C*). Pero para identificar esos otros conceptos, el funcionalista debe apelar a su vez a varias de las inferencias en las que ellos figuran. No podemos determinar si dos individuos comparten un concepto a menos que determinemos que comparten todas estas otras inferencias.

Ahora bien, dado el criterio antes especificado para delimitar una descripción funcional, identificaremos el contenido cognitivo de un concepto *C* con una descripción funcional de *C* que incluya sólo las inferencias por default en las que *C* aparece. Pero considero que esta estrategia no restringirá suficientemente el rol funcional de *C*. Pues aunque varios conceptos sean excluidos de la descripción funcional de *C*, los conceptos que quedan tienen que ser a su vez definidos funcionalmente (y, a su vez, también los que aparecen en la descripción de estos). Esto quiere decir que aunque el rol sea restringido, este sigue siendo definido en función de los roles (también restringidos) de otros conceptos, que a su vez son definidos por los roles de otros y así sucesivamente. La propuesta funcional requiere que para compartir un concepto no sólo compartamos algunas de las inferencias en las que aparece sino también algunas de las inferencias *en las que aparecen muchos otros conceptos*. Aunque la determinación del rol ya no es holista (ya no se requiere de *todas* las relaciones inferenciales), la lista de inferencias seguirá siendo suficientemente amplia como para hacer implausible la satisfacción del requisito de publicidad.

Este punto es importante. Fodor considera que los roles funcionales no permiten explicar la publicidad de un concepto dado *C* porque no tenemos un criterio (como la distinción analítico/sintético) para decidir, *entre las inferencias en las que figura C*, aquellas que constituyen su contenido. Sin embargo, lo recién dicho muestra que el problema es más grave. Aún si contamos con un criterio para restringir las inferencias en las que aparece *C* y *que son relevantes para individualarlo*, la individuación de *C* requiere también de muchas otras inferencias *en las que C no aparece*. El verdadero problema del holismo no se sigue sólo de la ausencia de una distinción analítico/sintético sino también del modo en que el funcionalismo debe individualar el rol funcional de un concepto.

Considero que una propuesta que integra a los roles funcionales pero que le da un lugar destacado a una teoría causal sí será capaz de satisfacer el requisito de publicidad. Antes de caracterizar mi propuesta debo hacer una estipulación terminológica. Llamaré “contenido causal amplio” al tipo de contenido que es determinado exclusivamente por algún tipo de relación con un estímulo distal, es decir, el tipo de contenido que postula una teoría causal. No emplearé la expresión “contenido causal” ya que, en su versión naturalista, el contenido funcionalmente entendido también involucra relaciones causales (aunque principalmente entre representaciones).

Según el enfoque que quiero proponer, entonces, hay dos elementos que constituyen el contenido de un concepto dado *C*: (i) La descripción del rol funcional de *C*, que nos permite identificar

un conjunto de representaciones $\{Q\}$ relacionadas inferencialmente con C , y (ii) alguna teoría causal, que nos permite identificar los contenidos causales amplios de las representaciones en $\{Q\}$. Diré que el contenido cognitivo de C no es el contenido causal amplio de C mismo, sino que es *el contenido causal amplio de las representaciones en $\{Q\}$* , el contenido causal amplio de las representaciones con las que C se relaciona inferencialmente. Pero esto quiere decir que si bien el contenido cognitivo de C no se identifica con su propio contenido causal amplio, sí se identifica con ciertos contenidos de este tipo. En este sentido la propuesta que ofrezco es una teoría causal del contenido cognitivo.

Por el contrario, no podemos decir estrictamente que sea una propuesta funcional. Si bien el rol funcional de un concepto es relevante para determinar su contenido cognitivo, este contenido no es identificado, ni en parte, con este rol funcional sino con los contenidos causales de las representaciones que aparecen la descripción de ese rol funcional. Mi propuesta no es entonces una propuesta funcional (pues el contenido no es identificado con una función) pero sí una propuesta causal (el contenido es identificado con *ciertos* contenidos causales). Así, si bien incorpora elementos de ambas propuestas no es una mera combinación de ambas. No incluye una tesis central del funcionalismo. Diré que sostengo una propuesta de contenido causal amplio *en* el rol funcional.

Así, para explicar los casos Frege emplearé una teoría causal compleja. Por un lado, el contenido cognitivo de un concepto C está determinado por los contenidos causales amplios de algunas de las representaciones que aparecen en su descripción funcional. Por otro lado, la referencia de C sí será identificada con su propio contenido causal amplio. Por ejemplo, tanto SUPERMAN como CLARK se relacionan causalmente con el mismo individuo y por lo tanto tienen la misma referencia. Sin embargo, están relacionados inferencialmente con representaciones que tienen diferentes contenidos causales amplios. SUPERMAN (pero no CLARK) está relacionado inferencialmente con una representación que se relaciona causalmente con la propiedad de volar y CLARK (pero no SUPERMAN) está relacionado inferencialmente con una representación que se relaciona causalmente con la propiedad de ser criado en Kansas. Esto hace que SUPERMAN y CLARK tengan contenidos cognitivos diferentes.

Veamos ahora cómo esta propuesta puede adoptar la solución vista al problema de la publicidad de los conceptos. Diré que el contenido cognitivo de un concepto C estará determinado sólo por los contenidos causales amplios de las representaciones que aparecen en las inferencias por default en las que figura C .

Vimos que apelar a inferencias por default no bastaba para que una propuesta de rol funcional resuelva el problema de la publicidad porque los roles así reducidos aún incluyen un conjunto de inferencias demasiado amplio. El problema era que la lista reducida de inferencias en las que entra un concepto C incluye conceptos que deben ser a su vez identificados por sus roles funcionales restringidos. A su vez, estos nuevos roles funcionales introducirán nuevos conceptos que deben ser definidos funcionalmente y así sucesivamente. En la propuesta que ofrezco aquí, el contenido de C es determinado también por los contenidos de un conjunto de representaciones $\{Q\}$ que aparece en la descripción funcional de C . Sin embargo, los contenidos de las representaciones en $\{Q\}$ no son, a su vez, definidos funcionalmente, sino por medio de una teoría causal. No se apela a su relación con otras representaciones sino con un estímulo distal. Esto hace que la lista relevante de inferencias no siga extendiéndose a pesar de haber reducido el rol funcional original (la lista de inferencias en las que aparece C). De este modo, es más plausible que la propuesta de contenido causal en rol funcional pueda satisfacer el requisito de publicidad adaptando la solución de Prinz.

5. Conclusión

En este capítulo presentado una explicación semántica de los casos Frege que incorpora tanto elementos de una teoría causal como elementos funcionales. El propósito de elaborar esta propuesta es superar los límites que estos dos tipos de teorías tienen para ofrecer una explicación de los casos Frege. Las teorías informacionales o generalmente causales no pueden explicar semánticamente los casos Frege porque son referencialistas. Las teorías de rol funcional no pueden hacerlo porque son holistas. La identificación de contenido cognitivo con la información o el contenido causal en el rol funcional permite que haya una dimensión informacional/causal del contenido que vaya más allá de la referencia y a la vez que el rol funcional no sea determinado holísticamente.

Capítulo 3:

Contenido cognitivo y forma sintáctica

1. Introducción

En el capítulo anterior mostré que Fodor motiva, por medio de un dilema, una explicación sintáctica de los casos Frege y que la posibilidad de esta explicación favorece una teoría informacional del contenido sobre una teoría de rol funcional. Argumenté que, sin embargo, es posible una explicación semántica de los casos Frege combinando ciertos elementos de las dos propuestas involucradas en el dilema. El resultado es una propuesta informacional (o, de modo general, causal) que involucra a los roles funcionales. Mostré que los roles funcionales permiten superar el referencialismo de las propuestas informacionales y que el enfoque informacional permite resolver el problema central que acarrea el holismo.

Esta línea de argumentación no dice nada en contra de la estrategia sintáctica misma o de propuestas informacionales alternativas para explicar los casos Frege. Sólo muestra que hay una explicación semántica e informacional o causal que involucra roles funcionales y que no acarrea los problemas atribuidos por Fodor a la información o a los roles funcionales. En este capítulo iré más lejos y argumentaré directamente contra la explicación sintáctica propuesta por Fodor, contra una explicación semántica informacional que ha surgido recientemente por parte de Jesse Prinz (2002) y en contra de la explicación informacional clásica de Dretske (1981).

2. La explicación sintáctica de los casos Frege

Retomemos brevemente el trasfondo de la discusión sobre los casos Frege que motiva la

explicación sintáctica de Fodor. Como he mostrado en el capítulo anterior, llamo “dilema de Fodor” a un problema que enfrentan las teorías del contenido mental para explicar los casos de Frege. Básicamente la idea es que la única alternativa relevante para explicar los casos Frege son las semánticas de rol funcional. Pero estas adolecen del problema del holismo semántico y por lo tanto son inaceptables. Por otro lado, las teorías causales, al ser atomistas, parecen ser las propuestas adecuadas para evitar el problema del holismo, pero como son referencialistas (identifican el contenido con la referencia) no cuentan con los recursos para explicar los casos Frege. Por lo tanto, si explicamos los casos Frege, caemos en el holismo. Pero si evitamos en el holismo, no seremos capaces de explicar los casos Frege.

Repasemos los elementos de este problema: Los casos Frege, la semántica de rol funcional y la semántica causal o informacional. Como vimos, los casos Frege son aquellos casos en los que la sustitución de expresiones de un lenguaje que tienen la misma referencia no preserva el valor de verdad de la fórmula de la que forman parte. Esto sucede principalmente en los llamados contextos lingüísticos opacos de sustitución. Estos incluyen, en particular, los contextos de atribución de actitudes proposicionales (como creencias o deseos). Por ejemplo, aunque sea verdadera la oración “Juan cree que la estrella vespertina es bella”, puede no ser verdadera la oración “Juan cree que la estrella matutina es bella”, aún cuando “estrella vespertina” y “estrella matutina” tengan la misma referencia. Aunque Frege haya planteado este problema en el marco de la caracterización de una noción no psicológica del contenido, vimos que se ha considerado frecuentemente que los casos Frege introducen restricciones para construir una semántica de las representaciones mentales. La explicación psicológica necesita que la semántica de la representación mental dé cuenta del hecho de que Juan puede creer que la estrella matutina es bella, puede tener esta actitud proposicional, sin creer que la estrella vespertina es bella (Fodor 2008, p. 52). Se ha considerado usualmente que lo que los casos Frege muestran es que *el contenido* de las representaciones mentales no puede ser identificado con la referencia. Si alguien puede tener la representación ESTRELLA MATUTINA sin tener la representación ESTRELLA VESPERTINA, entonces, dado que estas representaciones tienen la misma referencia, el contenido de las mismas no puede identificarse con aquella referencia.

He mencionado en el capítulo pasado que, según Fodor (2008), la alternativa más común a una teoría referencialista del significado (una que identifique el contenido con la referencia), es una teoría de dos factores donde el contenido sea caracterizado por medio de la referencia y otra dimensión semántica que puede llamarse “sentido”. Fodor considera que hoy en día los sentidos son ampliamente

identificados con roles funcionales y por lo tanto la teoría que puede dar cuenta de ellos es la llamada “semántica de rol funcional” (Fodor 2008, p. 53). Dicho muy brevemente, según esta propuesta, el contenido de una representación está determinado por las relaciones causales o inferenciales que tiene con otras representaciones. Ahora bien, vimos que Fodor ha argumentado extensamente que este tipo de propuestas adolece del problema del holismo (por ej., Fodor y Lepore 1992 y Fodor y Lepore 2002). Como he dicho previamente, el problema es que el criterio que necesitamos para determinar cuáles de las inferencias o relaciones causales en las que entra una representación determinan su contenido involucra la distinción analítico/sintético criticada por Quine (1953). Como no podemos admitir este criterio, debemos identificar al contenido de un concepto con su rol funcional total, con *todas* las inferencias en las que figura. Esto constituye un problema principalmente por dos razones. En primer lugar, si la semántica es holista en el sentido descripto, entonces dos personas que difieran en alguna de esas inferencias (o peor aún, una misma persona en diferentes momentos) no tendrán el mismo concepto. Pero es plausible que estas diferencias se den masivamente. Esto quiere decir que el requisito de publicidad para una teoría de conceptos, la exigencia de que los conceptos sean compartidos, no puede ser satisfecho por esta propuesta. Más aún, esto implica que la teoría de rol funcional es incompatible con generalizaciones psicológicas, con la formulación de las leyes intencionales que constituyen la explicación psicológica de la conducta, según Fodor (1994). Por otro lado, Fodor también ha argumentado que una semántica holista tampoco es apta para satisfacer el requisito de composicionalidad para una teoría de conceptos. Si los conceptos son identificados con roles funcionales totales, entonces los conceptos no se combinan composicionalmente, es decir, el significado de un concepto complejo no puede reducirse al significado de sus partes. El ejemplo clásico de Fodor es el caso en el que el concepto PELIGROSA forma parte del rol funcional del concepto complejo VACA MARRÓN sin que PELIGROSA forme parte del rol funcional de VACA o de MARRÓN (Fodor y Lepore 1992 y Fodor y Lepore 2002).

Como vimos, la propuesta alternativa que Fodor ha favorecido en diversas ocasiones es una semántica de tipo informacional. Para una primera y esquemática aproximación, las propuestas informacionales sostienen que una representación *R* representa a una propiedad *P* si y sólo si hay una ley que determina que los casos de *P* causen casos de *R* (Fodor 1998). Como señalé en el capítulo 2, Fodor (2008) no adhiere específicamente a una teoría informacional pero sí se compromete con la familia de teorías de la que esta teoría forma parte. Afirma que algún tipo de teoría causal es verdadera (Fodor 2008, p. 199) e incluso ocupa un capítulo entero (capítulo 7) a responder a las objeciones

tradicionales que se le han presentado a este tipo de enfoque. Estas propuestas tienen, según Fodor, dos ventajas cruciales. En primer lugar, parece ser una de las pocas teorías aptas para naturalizar el contenido mental. A pesar de que este es un aspecto muy relevante de las propuestas causales, no me detendré en él ya que no es indispensable para el problema que me ocupa. La segunda ventaja que Fodor atribuye a las teorías causales, y la que resulta relevante para el problema planteado, es su atomismo. Vimos que según esta propuesta, el contenido de una representación R de un sujeto S es determinado puramente por la relación causal que tiene con una propiedad distal. Esto quiere decir que las otras representaciones que S posea y las relaciones (inferenciales o causales) que tengan con R no afectan de ningún modo el contenido de R . En esto consiste el atomismo de la propuesta causal y es esta característica la hace inmune al holismo semántico. No tenemos que determinar qué inferencias son relevantes para determinar el contenido de una representación porque ninguna lo es.

Ahora bien, a pesar de estas ventajas, hay un aparente punto débil de las teorías causales. Este punto, junto con el fracaso de la semántica de rol funcional, es lo que motiva la explicación sintáctica de los casos Frege. Fodor sostiene que las propuestas causales son referencialistas. Esto quiere decir que, según este enfoque, las expresiones que tengan la misma referencia serán sinónimas, tendrán el mismo significado. Por ejemplo, si adoptamos una semántica informacional y si asumimos que H_2O y agua son la misma propiedad, entonces si el concepto AGUA está en una relación causal nomológica con agua, también lo estará con H_2O y si el concepto H_2O está en una relación causal nomológica con H_2O también lo estará con agua. Por lo tanto, los conceptos AGUA y H_2O serán sinónimos (Fodor 1998). Ahora bien, como vimos, se considera comúnmente que los casos de tipo Frege son contraejemplos para una semántica referencialista. El contenido no puede identificarse con la referencia.

Dada esta situación dialéctica, si los casos Frege requieren una explicación semántica, nos enfrentamos a un dilema. Por un lado, si adherimos a alguna teoría causal del contenido evitamos el holismo pero no podemos explicar los casos Frege. Por otro lado, si adherimos a una semántica de rol funcional podemos explicar los casos Frege pero no podemos evitar el holismo semántico.

La respuesta fodoriana a este dilema, a través de diferentes trabajos, (por ej., 1994, 1998, 2008) es la idea de que los casos de Frege no precisan la introducción de una nueva dimensión del contenido conceptual sino que pueden ser explicados a partir de propiedades no semánticas, puramente sintácticas de los conceptos. Fodor (2008) acepta que los casos Frege muestran que los conceptos no pueden ser

individuados sólo en base a la referencia, pero afirma de ello no se sigue que aquello adicional que precisamos para la individuación sea un parámetro del contenido. Fodor (2008) divide su explicación de los casos Frege en dos partes. En la primera considera el caso de los conceptos complejos y en la segunda se ocupa de los conceptos básicos. Los conceptos complejos son aquellos que poseen una estructura constituyente. Los constituyentes de una representación son sus partes semánticamente interpretables. La idea de Fodor es que dos conceptos que tienen la misma referencia pero diferentes constituyentes tendrán diferentes condiciones de posesión. Por ejemplo, uno puede tener el concepto ESTRELLA MATUTINA pero no ESTRELLA VESPERTINA si uno carece del concepto VESPERTINA, ya que la posesión de VESPERTINA es una condición para la posesión de ESTRELLA VESPERTINA pero no para la posesión de ESTRELLA MATUTINA. De este modo explicamos que alguien pueda poseer uno de dos conceptos que tienen la misma referencia sin poseer el otro. Esto es una explicación sintáctica de los casos Frege para los conceptos complejos porque los constituyentes, si bien son partes interpretables semánticamente, son constituyentes sintácticos (es decir, que pueden ser individuados sintácticamente, sin apelar a propiedades semánticas) (2008, p. 61).

Antes de considerar el caso de los conceptos simples, es importante volver a enfatizar un punto sobre los constituyentes los conceptos complejos que resulta crucial para varios puntos centrales de este capítulo. Cuando Fodor habla de constituyentes de un concepto complejo, está hablando de partes contenidas en el concepto, de su estructura interna. En particular, un concepto *R* sólo relacionado inferencialmente con otro *C* no es un constituyente de *C*. Por ejemplo, PERRO MARRÓN no es un concepto que tenga como constituyente a ANIMAL, aunque esté relacionado inferencialmente con él. Pero sí tiene como constituyente a PERRO (con el cual está además inferencialmente relacionado). Esto es importante para notar que la admisión de conceptos complejos no implica la negación del atomismo respecto de los conceptos. Si bien necesito tener PERRO y MARRÓN para tener el concepto PERRO MARRÓN, no necesito ningún otro concepto para tener PERRO, o cualquier otro concepto básico (o léxico), es decir, un concepto sin diferentes componentes sintácticos. El atomismo es una tesis sobre los conceptos básicos. Como veremos, este punto será relevante para argumentar en contra de la explicación sintáctica de los casos Frege y para entender las propuestas de Dretske y de Prinz, y la evaluación de Prinz de la propuesta fodoriana.

Ahora, para el caso de los conceptos básicos, aquellos que no tienen una estructura constituyente, también pueden darse los casos de Frege. Fodor considera que los mejores candidatos a ser conceptos básicos son los expresados por nombres monomorfémicos de individuos. Por ejemplo,

CICERÓN y TULIO pueden ser conceptos básicos que si bien tienen la misma referencia parece posible tener uno sin tener el otro. Si bien Fodor considera otros casos, no se compromete con otros tipos de conceptos simples.

Como hemos visto, Fodor considera que lo que los casos Frege motivan es la introducción de un criterio de individuación de conceptos que permita que conceptos con igual contenido sean diferentes conceptos. Necesitamos decir que dos conceptos son diferentes conceptos, según Fodor, *cuando tienen diferentes poderes causales*. Fodor supone que es conocimiento común que las causas y efectos de pensar sobre una cosa depende de cómo esa cosa sea conceptualizada. Por ejemplo las causas y efectos de pensar que Venus es húmedo dependen de si conceptualizamos a Venus como ESTRELLA MATUTINA o como ESTRELLA VESPERTINA (p. 61). Como hay casos en los que conceptos con la misma referencia *tienen diferentes poderes causales*, no podemos individuar a los conceptos sólo en base a la referencia. Por ejemplo, como creer que Cicerón era gordo puede causar que afirme “Cicerón era gordo” pero creer que Tulio era gordo no CICERÓN y TULIO tienen que ser diferentes conceptos, a pesar de tener la misma referencia.

La propuesta de Fodor es emplear la tesis de que los procesos mentales son computacionales (lo que llama “teoría computacional de la mente”) para proporcionar un criterio de individuación de conceptos que explique estos casos. La teoría computacional afirma, según Fodor, que el aspecto causalmente relevante de un concepto (o de la fórmula del lenguaje del pensamiento de la que forma parte) no es su contenido sino su sintaxis. Si esto es así, *se explica que conceptos con el mismo contenido puedan tener poderes causales diferentes*. Simplemente tendrán propiedades sintácticas diferentes. Pero esta diferencia sintáctica no es una diferencia en constituyentes, ya que se trata de conceptos básicos, sino de *forma*. La forma sintáctica de una representación es simplemente alguna propiedad física (a Fodor no le interesa determinar específicamente qué es) a la que es sensible la computación mental (o neuronal).

Ahora bien, considero que noción de sintaxis no sólo no constituye una solución al dilema sino que nos revela que ni siquiera hay un dilema. En el capítulo anterior argumenté que una tesis central de un cuerno del dilema, la de que (i) los roles funcionales no son composicionales, depende de un argumento que presupone la tesis de que (ii) la forma sintáctica constituye un criterio de individuación para los conceptos. Esto era un problema porque (ii) se basa, en parte, en (i) y el argumento a favor de (i) no funciona sin presuponer (ii). En este capítulo quiero argumentar que, aún si obviamos la

circularidad de aceptar (ii) para argumentar a favor de (i), la noción de forma sintáctica fodoriana mina el argumento a favor de (i) y por lo tanto, el dilema que nos lleva a la aceptación de (ii).

Empecemos recordando brevemente el requisito de composicionalidad. Fodor (2008) no desarrolla el problema que los roles funcionales tienen con el requisito de composicionalidad sino que simplemente nos remite a Fodor y Lepore (2002). Fodor y Lepore (2002) afirman que la composicionalidad es una propiedad que un sistema de representaciones tiene cuando (i) contiene símbolos primitivos y símbolos que son sintáctica y semánticamente complejos, y (ii) los últimos heredan sus propiedades semánticas y sintácticas de los primeros (2002, p.1). Vimos que el argumento central de Fodor en contra de los roles funcionales parte del holismo de los roles (dado por la ausencia de distinción analítico/sintético) y muestra que dado este holismo y consideramos lo que llame “casos Fodor”, los roles funcionales no se combinan composicionalmente.

Ahora bien, considero que los casos Fodor dependen de una caracterización de los conceptos complejos que es inconsistente con la noción fodoriana de forma sintáctica. Los casos Fodor presuponen que los conceptos complejos tienen una semántica caracterizada funcionalmente. Esto es así porque, como vimos, estos casos funcionan en el contexto de una reducción al absurdo del funcionalismo. Fodor pretende mostrar que si suponemos que los contenidos conceptuales son roles funcionales entonces los conceptos no se combinan composicionalmente. Pero como sabemos que los conceptos sí se componen de este modo, podemos concluir que los contenidos conceptuales no son roles funcionales.

Además, como señalé también en el capítulo anterior, los casos Fodor suponen que parte de las condiciones de individuación de un concepto complejo es su estructura compleja *sintáctica*. En el caso de Fodor, el concepto que representa a las vacas marrones tiene un rol funcional que incluye al rol de VACA y al de MARRÓN pero que no se reduce a ellos. Vimos que esto es un problema solamente porque independientemente de este contenido identificado funcionalmente, sabemos que el concepto está compuesto *sólo* por VACA y por MARRÓN. Pero para poder afirmar que estos son sus únicos componentes no podemos basarnos obviamente en el rol funcional del concepto (puesto que este involucra algo más que los roles de VACA y MARRÓN) sino en algún otro aspecto suyo. Teniendo el trasfondo de la propuesta fodoriana, el único candidato relevante parece ser su estructura sintáctica. El concepto que refiere a las vacas marrones tiene una estructura interna que tiene como componentes *sintácticos* sólo a VACA y a MARRÓN y *un rol funcional no se reduce al de estos dos conceptos*.

Ahora bien, considero que la caracterización fodoriana de la noción de forma sintáctica es incompatible con que se den los casos Fodor así caracterizados. Como vimos más arriba, la forma sintáctica de un concepto es para Fodor aquella que *determina los poderes causales de una representación*. La forma de VACA es la que es responsable de que este concepto cause, por ejemplo, instancias de ANIMAL. Pero si esto es así, no se explica de qué modo un concepto que tiene como componentes *sintácticos* sólo a VACA y a MARRÓN tiene poderes causales que no posee ni VACA ni MARRÓN. Si un concepto tiene estos poderes adicionales, parece que debemos asumir que hay algo más en su estructura sintáctica. Pero si esto es así, el hecho de que el rol funcional de este concepto no se reduzca a los roles de VACA y MARRÓN no implica una violación de composicionalidad. La parte adicional del rol funcional puede ser atribuida a este componente sintáctico adicional.

Uno podría intentar responder a este argumento afirmando lo siguiente: Aunque el concepto que representa a las vacas marrones sólo está compuesto por VACA y por MARRÓN, puede haber alguna propiedad sintáctica que no está en VACA ni en MARRÓN aisladamente pero que resulta de su combinación. Esto puede explicar los poderes causales adicionales del concepto complejo sin necesidad de agregar otro *concepto* en su composición. Pero esta estrategia no puede funcionar. Pues como hemos visto, el requisito de composicionalidad no se aplica, según Fodor y Lepore (2002) meramente a las propiedades semánticas sino también a las sintácticas. De este modo, para que se cumpla composicionalidad, las propiedades sintácticas de un concepto complejo deben poder reducirse a las propiedades sintácticas de sus componentes.

Supongamos que podemos evitar este problema, que podemos mostrar cómo un componente sintáctico adicional, uno que no pertenece a ninguno de dos conceptos, puede surgir la combinación de ellos y que esto no implique una violación de composicionalidad. Aún si esto es así, tenemos otro problema. Si un componente sintáctico adicional surge *de la mera combinación* de, por ejemplo, de VACA y MARRÓN, y este explica que VACA MARRÓN cause PELIGROSA, entonces no podemos formar VACA MARRÓN sin que PELIGROSA esté incluido en su rol funcional. Pero claramente este no es el caso. PELIGROSA y muchos otros conceptos llegan a estar asociados a un dado *C* no como resultado de la formación misma de *C* sino como resultado de encuentros con ejemplares de la categoría que *C* representa. No aprendo por la mera combinación de VACA y MARRÓN que las vacas marrones son peligrosas.

He argumentado, entonces, que si nos tomamos en serio la noción de forma sintáctica, los casos

Fodor no pueden darse. Ahora bien, los casos Fodor funcionan en contra de la composicionalidad de los roles funcionales, pero esta no es el único problema que Fodor atribuye a los roles funcionales. Como vimos, considera también que el holismo de los roles funcionales hace imposible que satisfagan el requisito de publicidad. Si Fodor puede mostrar que la forma sintáctica puede explicar los casos Frege sin caer en ese mismo problema, entonces aún tenemos una motivación para adoptar la estrategia sintáctica.

Veamos brevemente la estrategia de Fodor (2008) para lidiar con el requisito de publicidad. Vimos que Fodor explica los casos Frege sosteniendo que las condiciones de identidad de un concepto no incluyen sólo su referencia o contenido informacional sino también su forma sintáctica. Ahora bien, Fodor nota que estas condiciones de individuación resultan problemáticas cuando queremos determinar si dos conceptos *en diferentes individuos* son del mismo tipo. Vimos que la forma sintáctica es una propiedad física, presumiblemente neurológica, de un concepto, que determina sus poderes causales. Pero no parece plausible suponer que el mismo concepto básico es realizado por la misma propiedad neurológica en diferentes mentes. Fodor considera que mientras más criaturas tengan una misma actitud proposicional, menos plausible es que esa actitud proposicional sea implementada del mismo modo en todas las criaturas que la poseen. Pero por otro lado, la generalidad de la psicología intencional es una virtud epistémica de la misma. Mientras se aplique a más criaturas entonces tenemos más razones para pensar que es cierta. Surge entonces el problema de cómo conciliar la generalidad de la psicología intencional con el hecho de que las formas sintácticas de las que parece precisar no son compartidas por los individuos a los que se aplica.

La solución fodoriana es sostener que no sólo los contenidos intencionales *sino también las propiedades computacionales, como la forma sintáctica*, pueden realizarse en diferentes bases neuronales (2008, p. 90). Como afirma Fodor, es más o menos conocimiento común que existe la posibilidad, al menos en principio, que los estados computacionales sean realizados múltiplemente en diferentes cerebros. Casos de la misma representación mental pueden ser realizados por propiedades sub-computacionales diferentes que son del mismo tipo bajo descripción computacional. Es tan implausible, según Fodor, que las explicaciones computacionales proporcionen reducciones de las explicaciones intencionales como que las explicaciones neurológicas proporcionen reducciones de las explicaciones computacionales.

De este modo por un lado la posibilidad de realización de un mismo estado intencional en

diferentes estados computacionales permite explicar los casos Frege. Por otro lado la posibilidad de realización de un mismo estado computacional en diferentes estados neuronales permite que los conceptos, individuados en base al contenido y la forma sintáctica, sean compartidos por diferentes individuos.

Ahora bien, nuevamente, considero que la noción fodoriana de forma sintáctica no permite esta solución al problema de la publicidad. En particular, creo que la publicidad de la forma sintáctica es paralela a la de los roles funcionales. Como dije antes, según Fodor, la forma sintáctica es lo que determina los poderes causales de una representación. Esto quiere decir que *cualquier* diferencia en los poderes causales de dos representaciones debe ser explicada por una diferencia en su forma sintáctica. Ahora bien, como hemos visto, los roles funcionales no son compartidos porque es muy plausible que siempre haya diferencias, aunque sean mínimas, entre las relaciones causales que tiene una representación R en un individuo S con otras representaciones en S y las relaciones causales que tiene otra representación R_1 (con el mismo contenido intencional que R) en otro individuo S_1 , con otras representaciones en S_1 . Pero si esto es así y si las relaciones causales de una representaciones están determinadas por su forma sintáctica, las formas sintácticas tampoco serán compartidas. Si las relaciones causales dependen de las formas sintácticas y aquellas no son compartidas, estas no lo serán tampoco. Entonces, aún si los estados computacionales pueden ser realizados en diferentes estados neuronales, si los roles causales no son compartidos, tampoco lo serán las formas sintácticas.

Podemos ahora concluir que no tenemos una motivación adecuada para una explicación sintáctica de los casos Frege. Pues por un lado, la forma sintáctica no permite que se den los casos Fodor, que son la base de uno de los argumentos en contra de la explicación semántica y a favor de la explicación sintáctica. Por otro lado, el otro argumento en contra de la explicación semántica, el hecho de que los roles funcionales no sean compartidos, implica que las formas sintácticas tampoco serán públicas.

Habiendo argumentado en contra de la explicación sintáctica de los casos Frege, en las dos secciones que siguen quiero considerar dos de las explicaciones semánticas e informacionales que se han propuesto para estos casos. En primer lugar, consideraré la explicación clásica de Dretske (1981) uno de los padrinos del enfoque informacional. En segundo lugar, consideraré la explicación más reciente de Jesse Prinz (2002, 2005).

3. Información digital y conceptos primitivos

Dretske (1981) afirma que si bien la noción de información nos sirve para elucidar la noción de contenido mental, no todo estado informacional cuenta como teniendo un contenido de este tipo. Dretske considera que la nota principal de la intencionalidad es el tipo de opacidad semántica que parecen revelar los casos Frege. Es decir, un estado posee contenido intencional si puede referirse a una de dos propiedades coextensivas sin referirse a la otra. Por ejemplo, un estado I es intencional si tiene como contenido a una propiedad P pero no a otra Q a pesar de que no haya nada que sea P y no sea Q .

Para Dretske, todo estado informacional es intencional en este sentido, pero no todo estado informacional es cognitivo. Esto se explica por el hecho de que, según el autor, la intencionalidad tiene grados. Todo estado informacional tiene un primer grado de intencionalidad, pero sólo los estados cognitivos son estados informacionales que poseen grados superiores de intencionalidad. Veamos primero en que consisten estos grados.

Un estado E posee un primer nivel de intencionalidad cuando se cumple que (i) Todos los F son G , (ii) E tiene el contenido de que t es F y (iii) E no tiene el contenido de que t es G . Por ejemplo, aunque todos los hijos de Juan tienen sarampión, y E porta la información de que t es uno de los hijos de Juan, E puede no portar la información de que t tiene sarampión.

A pesar de que Dretske no lo explicita aquí, esto puede ocurrir cuando a pesar de que las propiedades son coextensivas, no están relacionadas nómicamente. Es nómicamente posible que algo sea el hijo de Juan y no tenga sarampión. La información depende de una relación contrafáctica o nómica entre el estado que porta información y su fuente. E está relacionado nómicamente con la propiedad de ser hijo de Juan, pero como esta propiedad no está relacionada nómicamente con la de tener sarampión, entonces E puede no estar tampoco relacionado nómicamente con tener sarampión. Todos los estados informacionales pueden portar este nivel de intencionalidad.

El segundo nivel de intencionalidad se da cuando un estado E cumple con las condiciones de que (i) hay una ley que determina que todos los F son G , (ii) E tiene el contenido de que t es F y (iii) E no tiene el contenido de que t es G . Por ejemplo, uno puede creer que el agua se está congelando sin creer que se está expandiendo aunque es nómicamente imposible que el agua se congele sin expandirse. Por último, el tercer nivel de intencionalidad se da cuando se cumple, para un estado E que (i) es

analítica o conceptualmente necesario que los F son G , (ii) E tiene el contenido de que t es F y (iii) E no tiene el contenido de que t es G . Por ejemplo podemos saber que la solución a una ecuación es 23 sin saber que la solución a ese problema es la raíz cúbica de 12,167 aunque sea conceptualmente imposible que algo sea 23 pero no la raíz cúbica de 12,167.

Según Dretske, los estados cognitivos exhiben un genuino contenido semántico porque, a diferencia de otros estados, poseen estos segundos y tercer niveles de intencionalidad. Por el contrario, el contenido informacional carece de los dos niveles superiores de intencionalidad. Si hay una ley nómica o conceptual que vincula a los F con los G entonces lo que porte la información de que algo es F necesariamente portará la información de que eso es un G . Tanto el contenido informacional como el contenido semántico de E son proposicionales, ambos consisten en el contenido de que t es F , pero sólo el contenido semántico de E consiste *exclusivamente* en el contenido de que t es F .

Ahora bien, Dretske cree que a pesar de que el contenido semántico no es el contenido informacional, puede ofrecer una explicación informacional del contenido semántico si explica cómo los contenidos semánticos *surgen a partir de los contenidos informacionales* (p. 175). Para este propósito Dretske retoma la distinción que propuso en el capítulo anterior para distinguir entre representaciones perceptuales y conceptuales: La distinción entre codificación analógica y digital de la información. Empecemos entonces explicando brevemente en qué consisten estos dos tipos de codificación de la información semántica.

El modo común de pensar la distinción entre la codificación digital y analógica es como la distinción entre un modo discreto o uno continuo de representar una propiedad variable en una fuente. Por ejemplo, el velocímetro de un auto constituye una representación analógica de la velocidad de un auto porque las diferentes velocidades son representadas por las diferentes posiciones de la aguja del velocímetro y estas posiciones varían de forma continua y cada posición representa un valor diferente de la propiedad representada. Por otro lado, la luz que registra la presión de aceite del auto es un dispositivo digital, ya que tiene dos estados (encendido y apagado) y no tiene estados intermedios informacionalmente relevantes. Dretske quiere usar una distinción análoga a esta pero aplicada no a la representación del conjunto de valores de una propiedad variable sino a la representación de la instanciación de determinada propiedad. Por ejemplo, no le interesa mostrar cómo podemos representar la temperatura sino el hecho de que la temperatura está a 100 grados centígrados, de que está demasiado elevada. Quiere mostrar que hay una distinción semejante a analógico-digital para

representar hechos (por ejemplo, el hecho de que s es F).

Dretske afirma que una señal porta la información de que s es F de modo digital si s no porta información adicional sobre s , información que no esté ya anidada en el hecho de que s es F . Si una señal porta información adicional sobre s información que no está anidada en el hecho de que s es F entonces porta la información de que s es F de forma analógica. Cuando porta la información de que s es F de forma analógica, una señal porta información más específica, más determinada que la información de que s es F . Toda señal porta información de forma digital y de forma analógica. La información más específica que una señal porta sobre s es la única pieza de información que porta en forma digital. Toda otra pieza de información que porte sobre s es información en forma analógica. Para ilustrar la diferencia entre estos dos modos de codificación Dretske toma el caso de una imagen y un enunciado. Supongamos que la taza tiene café y que yo quiero comunicar esta información. Si yo simplemente digo “la taza tiene café”, esta señal acústica porta la información de que la taza tiene café en forma digital. No porta información más específica que la de que la taza tiene café (cuánto café hay, cuán oscuro es, cuán grande es la taza, etc.). En cambio, si fotografío la taza y la muestro, la fotografía porta la información de que la taza tiene café en forma digital. La foto dice que hay café en la taza diciendo, vagamente, cuánto café hay en la taza, diciendo el color, tamaño y forma de la taza, etc.

Dretske (1981) emplea esta distinción para explicar los procesos cognitivos y señalar una diferencia entre representaciones perceptuales y conceptuales. Según Dretske, la esencia de la cognición es la conversión digital. Los procesos cognitivos son procesos que extraen información contenida de modo analógico en nuestros sistemas perceptuales para convertirla en información codificada digitalmente. Los conceptos son básicamente los medios por los que llevamos a cabo tareas como la generalización, clasificación y el reconocimiento y todos estos procesos requieren ignorar mucha de la información específica contenida en nuestra experiencia y recoger información más general, común a diferentes experiencias perceptuales (cap. 6).

Volvamos ahora a la explicación informacional del contenido semántico. La idea de Dretske es identificar al contenido semántico de una estructura E con la información que porta sobre un objeto o en forma digital, la información más específica que porta sobre o . Por ejemplo, si una señal s porta la información de que t es un cuadrado pero no porta información más específica sobre t (por ejemplo, que es rojo o azul, grande o pequeño, etc.) s porta la información de que t es un cuadrado en forma digital y por lo tanto la tiene como contenido semántico. Al portar esta pieza de información en forma

digital s porta mucha otra información en forma analógica (por ejemplo, la información de que t es un cuadrilátero, que t es un paralelogramo, que t es un rectángulo). Estas son piezas de información que están anidadas en la información de que t es un cuadrado, pero que no constituyen el contenido semántico de s . Dretske considera que la información en forma digital posee los niveles de intencionalidad requeridos para los estados cognitivos como las creencias. Si s porta la información de que t es F la información de que t es G está anidada (nómica o conceptualmente) en la información de que t es F , es decir, hay una ley que determina que todo F es G o es una verdad conceptual que todo F es G , entonces se sigue que s porta la información de que t es G . s no puede tener una pieza de información como su contenido informacional sin que la otra también lo sea. Pero s sólo tiene la información de que t es F como contenido semántico porque sólo porta esta información en forma digital. La señal s cumple con las condiciones para que su contenido posea los dos niveles superiores de intencionalidad que Dretske atribuye a los estados cognitivos (cap. 7).

Ahora bien, considero que esta caracterización del contenido semántico en términos de información digitalizada no es adecuada. Creo que la noción de información digitalizada no rescata adecuadamente la opacidad semántica propia de los estados cognitivos. Los casos Frege muestran que las condiciones de individuación de un concepto permiten que un sujeto tenga uno de dos conceptos *correferenciales* sin tener el otro. Por ejemplo, alguien puede creer que la estrella matutina es bella sin creer que la estrella vespertina es bella aunque los conceptos ESTRELLA MATUTINA y ESTRELLA VESPERTINA que forman parte de estas dos creencias tengan la misma referencia, refieran al mismo individuo o conjunto de individuos. Los casos Frege, como vimos repetidamente, muestran que la individuación de conceptos involucra algo más que la referencia.

Por el contrario, la versión de los casos Frege que obtenemos a partir del enfoque de Dretske, no conserva estas características centrales. Los casos Frege requieren dos conceptos diferentes que comparten cierta parte del contenido pero difieren en otro respecto. Dado el marco de Dretske un caso Frege se daría cuando tenemos dos conceptos C y C_1 tales que *ambos* portan la información de que t es G pero C sólo porta la información de que t es F en forma digital y por lo tanto la tiene como su contenido semántico, mientras que C_1 porta sólo a la información de que t es G en forma digital y por lo tanto la tiene como su contenido semántico. Ambos conceptos comparten un contenido informacional (que t es G) pero tienen contenidos semánticos diferentes. Como en los casos Frege tenemos el caso de dos conceptos que comparten en cierto aspecto de su contenido pero difieren en otro aspecto. Sin embargo, el aspecto en el que difieren, el contenido semántico *no es otra cosa que la*

referencia. La propiedad G que C_1 tiene como contenido semántico es portada de forma analógica por C . Como vimos, esto quiere decir que G es una propiedad más general que F (la propiedad que C tiene como contenido semántico). Es decir, es una propiedad que se aplica a más individuos, que tiene una extensión mayor. Por lo tanto, los casos Frege dretskeanos, no son genuinos casos Frege, no son casos en los que hay dos conceptos diferentes *aunque correferenciales*. Esto quiere decir que Dretske no nos proporciona, por medio de la noción de información digital, un criterio para distinguir entre conceptos correferenciales. Pero es justamente este criterio el que define la característica opacidad semántica de los estados cognitivos como las creencias.

Ahora bien, aunque la caracterización central de la intencionalidad es la basada en la noción de información digital, Dretske dice algo más sobre las condiciones de individuación de conceptos. En el capítulo 9, Dretske (1981) reconoce que una teoría de conceptos debe proporcionar un criterio para distinguir conceptos co-referenciales y que la propuesta que él ha dado en el capítulo 7 no permite hacerlo. Si F y G son propiedades que son nómica o conceptualmente equivalentes (hay una ley empírica o una verdad conceptual que determina que F si y sólo si G) entonces una señal que constituya una digitalización de que t es F es una digitalización de que t es G y viceversa. Estas dos piezas de información digitalizadas, estos contenidos semánticos son indistinguibles desde el enfoque propuesto por Dretske. Sin embargo, admite Dretske, los conceptos que portan estos contenidos deben poder ser distinguidos (p. 215). Como solución a este problema Dretske admite que la opacidad semántica que el enfoque de Dretske explica es la que concierne a conceptos básicos, conceptos que no pueden descomponerse en estructuras cognitivas más simples. Dretske considera que el único modo en que un sistema puede distinguir dos conceptos que son nómica o conceptualmente equivalentes es si al menos uno de ellos es complejo, si uno está compuesto de elementos de los que el otro no. Por ejemplo si G y F son nómica o conceptualmente equivalentes, pueden ser representados por conceptos diferentes si sólo el concepto que tiene a G como contenido está compuesto por dos conceptos que tienen, respectivamente, a H e I como contenidos semánticos. A pesar de que estos conceptos tienen el mismo contenido semántico, son diferentes porque al tener una estructura interna distinta tener uno y no el otro conlleva diferencias conductuales. Por ejemplo, el sujeto que aplica uno de estos conceptos a t , pero no el que aplica el otro, tendrá la disposición a inferir que t es H .

Lo que Dretske afirma que es imposible, según su propuesta, es que dos conceptos primitivos y equivalentes sean diferentes. Si son equivalentes, tienen el mismo contenido y si son primitivos, es este contenido no articulado lo que determinará sus poderes causales. Una distinción entre estos dos tipos

de conceptos es, según Dretske, una distinción sin diferencia (p. 217). A partir de este punto, Dretske extrae a lo largo del capítulo 9 varias e importantes consecuencias respecto de la adquisición de conceptos y la adquisición de conocimiento.

Hay varios puntos que quiero señalar a partir de estas consideraciones de Dretske. En primer lugar, es importante notar que Dretske, al igual que Fodor, da explicaciones diferentes para los casos Frege que involucran conceptos primitivos y para los que involucran conceptos complejos. Como vimos, los casos de conceptos primitivos que el enfoque de Dretske explica, al no ser casos de conceptos extensionalmente equivalentes, no son genuinos casos Frege. Por lo tanto, como Dretske afirma que no hay distinción entre conceptos primitivos *extensionalmente equivalentes*, podemos concluir que Dretske no dispone de una explicación de los casos Frege para conceptos primitivos. A diferencia de la información digital, vimos que la forma sintáctica fodoriana sí nos permitía distinguir entre conceptos primitivos que correferían. Fodor muestra que, a pesar de lo que afirma Dretske, puede haber conceptos primitivos correferenciales que tengan diferentes consecuencias causales. Sin embargo, como he mostrado que no es admisible la forma sintáctica como criterio de individuación de conceptos, considero que es cierto, como creía Dretske, que no contamos con un criterio para distinguir entre conceptos primitivos equivalentes. Pero, como señalé, esto equivale a afirmar, en contra de Dretske, que no contamos con una explicación de la opacidad semántica de los conceptos primitivos.

En segundo lugar, la explicación de los casos Frege para el caso de los conceptos complejos es semejante a la ofrecida por Fodor. Los conceptos complejos equivalentes pueden distinguirse por su estructura constituyente. En la próxima sección argumentaré que los casos Frege tampoco pueden explicarse en base a la estructura interna de los conceptos.

4. La estructura interna semántica

Prinz (2002) considera que el problema de explicar los casos Frege dentro de un enfoque informacional puede ser resuelto si entendemos a los conceptos como *proxitypos*. Los proxitypos son mecanismos que poseen una estructura interna compleja, que pertenecen a sistemas perceptuales y que pueden ser usados por esos sistemas para detectar propiedades. Prinz considera que la capacidad de detección de los proxitypos les da a los conceptos las propiedades informacionales que necesitan para

dar cuenta del contenido intencional, mientras que la complejidad de la estructura permite explicar semánticamente los casos Frege, determinando para los proxytipos un contenido cognitivo.

Prinz considera que la idea de que los conceptos son estructurados contrasta con el atomismo de Fodor. Considera que, dado que los conceptos son estructurados, su posición es no atomista. Esto es relevante porque, como vimos, es la propuesta funcional la que típicamente Fodor caracteriza como no atomista. Sin embargo, considero que la propuesta de Prinz no es funcional. Para ver por qué, revisemos con un poco de detalle el tratamiento que le da Prinz a los casos Frege.

Los proxytipos son estructuras almacenadas en la memoria de largo plazo que pueden ser empleadas como detectores cuando son activadas en la memoria de trabajo. En tanto detectores, son estructuras que pueden covariar nomológicamente con individuos o con miembros de una clase. Esta relación de covariancia con una propiedad dada permite, según la semántica informacional, que refieran a dicha propiedad. Pero además (a diferencia de lo que Prinz llama “indicadores”) estas estructuras son complejas: están compuestas de representaciones de diferentes propiedades. Los proxytipos detectan alguna propiedad *P* que tienen como referencia detectando otro conjunto {*Q*} de propiedades superficiales o de apariencia que constituirán su contenido cognitivo (o, en terminología de Prinz, su contenido nominal). Esto quiere decir que proxytipos con la misma referencia pueden tener diferentes contenidos. Por ejemplo, CLARK puede ser un detector compuesto de ANTEOJOS, TRAJE MARRÓN, PEINADO HACIA ATRÁS, etc., mientras que SUPERMAN es un detector compuesto de CAPA ROJA, TRAJE AZUL, JOPO, etc. Son estructuras complejas que covarían con el mismo individuo (a quien por lo tanto refieren) covariando con diferentes conjuntos de propiedades suyas. De este modo, los proxytipos, siendo estructuras no atómicas, pueden ofrecer una explicación puramente informacional de los casos Frege.

Lo que me interesa antes que nada es mostrar por qué esta propuesta no constituye un anti atomismo en el sentido en que el funcionalismo lo es. Vimos que según Fodor las propuestas informacionales (y cualquier otra propuesta causal) son atomistas porque el contenido es puramente determinado por una relación con un objeto distal. Las relaciones entre representaciones quedan afuera de la determinación del contenido. Por lo tanto, para poseer un concepto no necesitamos poseer ningún otro. Como también vimos, esto sólo se aplica a los conceptos básicos o léxicos (aquellos que no tienen una estructura sintáctica interna). Para tener un concepto complejo sí necesitamos tener otros conceptos (necesitamos tener sus conceptos constituyentes). Ahora bien, Prinz se considera no atomista en el

sentido de que todos los conceptos tienen una estructura *interna* compleja en este sentido. Los conceptos, en tanto proxytipos, están compuestos por otras representaciones. Pero estas otras representaciones no son representaciones que se encuentren en relación inferencial o causal con el proxytipo en cuestión sino que simplemente son sus partes constituyentes. Los proxytipos son conjuntos de estructuras que se activan *al mismo tiempo*, se co-activan en la memoria de trabajo.

Es importante señalar que, a pesar de lo dicho, Prinz (2002) sí le da un lugar importante a las relaciones entre representaciones. Afirma que el modo en que podemos asociar representaciones entre sí es por medio de la formación de lo que llama “enlaces de memoria” entre diferentes proxytipos. Un conjunto de proxytipos unidos por enlaces de memoria constituye una red de memoria a largo plazo. Una red de representaciones puede contener información muy diversa sobre una categoría dada. Sin embargo, y esto es lo importante, los enlaces no desempeñan ningún papel en la determinación del *contenido* de un concepto. Para Prinz, lo que podemos decir sobre el contenido cognitivo e intencional de un concepto está plenamente determinado por relaciones informacionales o etiológicas. Estas condiciones determinan el contenido de una estructura por medio de las relaciones de covarianza o relaciones etiológicas que guarda con un objeto distal, pero no con otras representaciones. Así, la teoría semántica de Prinz no da lugar a la posibilidad de que el enlace que una representación dada tenga con otra pueda modificar su contenido.

He mostrado entonces que la propuesta de Prinz no involucra roles funcionales sino estructuras internas complejas. Ahora bien, he mostrado que no tenemos motivación para explicar los casos Frege a partir de estructuras sintácticas. Esto incluye tanto a las formas sintácticas como a las estructuras sintácticas complejas. Sin embargo, como vimos, Prinz explica los casos Frege a partir de una estructura interna compleja *semánticamente* caracterizada. El concepto CLARK y el concepto SUPERMAN difieren en cuanto al contenido informacional de sus componentes, los conjuntos de propiedades distales con los que covarían. Esto es relevante ya que, en la sección anterior, mostré que aunque no tenemos un criterio *informacional* para distinguir conceptos simples correferenciales, como sugirió Dretske, podríamos tenerlo para los conceptos complejos correferenciales. Esto es justamente lo que propone Prinz. A continuación argumentaré que, sin embargo, la propuesta de Prinz no constituye una explicación adecuada de los casos Frege.

Vimos que una de las consideraciones fregeanas que motiva la idea de un contenido cognitivo es que puede haber conceptos que sean diferentes a pesar de tener la misma referencia. Por ejemplo,

cuando le atribuyo a Louis Lane la creencia de que Clark Kent ha llegado no puedo inferir de ello la verdad de la atribución a Louis de la creencia de que Superman ha llegado, si que Louis desconoce la identidad entre Clark y Superman. Los conceptos CLARK y SUPERMAN de Louis tienen la misma referencia pero son diferentes. Tener una creencia que incluye uno de ellos no implica tener una creencia que incluye al otro. Sin embargo, y esto es lo que me interesará aquí, esta situación puede cambiar. Si Louis descubre la identidad del superhéroe, entonces los dos conceptos pasarán a ser equivalentes. Cuando Louis conoce la identidad del superhéroe entonces si sabemos que Louis cree que Clark ha llegado podemos inferir con verdad que Louis cree que Superman ha llegado. La razón para afirmar esto es que, como señaló Fodor (2008), cuando dos conceptos tienen la misma referencia, la motivación para pensar que son diferentes *es que tienen diferentes poderes causales*. Sin embargo, una vez que Louis conoce la identidad del superhéroe, *ya no tendremos una diferencia de este tipo*. Como Louis sabrá, por ejemplo, que Clark vuela y que Superman trabaja en el Daily Planet, no habrá diferencias conductuales frente al mismo individuo en las situaciones diferentes en las que tenga apariencia diferente. Llamaré a esta característica de un concepto, a la variación de las características involucradas en un caso Frege, “variación fregeana”. No hablo de “variación sintáctica” o “variación cognitiva” para no comprometerme con una lectura sintáctica o semántica de lo que exigen los casos Frege. Considero que toda explicación de los casos Frege debe poder dar cuenta de la variación fregeana.

Antes de pasar a considerar si la propuesta de Prinz puede dar cuenta de este fenómeno, es importante enfatizar una característica de los proxytipos. Los proxytipos son representaciones perceptuales. Esto quiere decir para Prinz que pertenecen a sistemas dedicados al procesamiento de un tipo de input específico (2002, cap. 5). Decir que los proxytipos pertenecen a estos sistemas implica que pueden ser activados por la presencia del input de ese sistema, esto es, que su activación *puede ser parte de la respuesta del sistema a ese input*. Cuando esto ocurre, los proxytipos funcionan como detectores perceptuales de estímulos distales.

Ahora bien, si Louis tiene un proxytipo que usa para detectar perceptualmente a Clark Kent (una estructura, digamos, compuesta por ANTEOJOS, TRAJE MARRÓN, PELO HACIA ATRÁS, etc.) en un momento t_0 y otro diferente que usa para detectar perceptualmente a Superman (uno compuesto, por ejemplo, de CAPA, TRAJE AZUL, JOPO, etc.) en un momento t_1 , Prinz tendrá que atribuir a Louis dos conceptos con contenido cognitivo diferente en t_0 y t_1 . Sin embargo, es claro que aún después de que Louis sepa de la identidad del superhéroe, usará esos proxytipos diferentes para

detectar perceptualmente al mismo individuo (pues dicho individuo seguirá teniendo una apariencia totalmente diferente en diferentes situaciones). Pero, como vimos, no podemos seguir atribuyendo a Louis en esas situaciones conceptos semántica o sintácticamente diferentes. Vimos que si Louis descubre la identidad del superhéroe, entonces los dos conceptos CLARK y SUPERMAN pasarán a ser cognitiva o sintácticamente equivalentes. Siempre podremos sustituir uno por el otro en contextos intensionales de atribución cuando Louis sea el sujeto. De este modo, el concepto atribuido en t_0 y t_1 no puede identificarse con las estructuras empleadas como detectores perceptuales en esos momentos. La explicación que demos de los casos Frege debe ser capaz no sólo de dar cuenta de la diferencia entre dos conceptos extensionalmente equivalentes, sino de la posibilidad de que esa diferencia se pierda, dada una nueva situación epistémica o doxástica de un sujeto. Es decir, debe dar cuenta de la variación fregeana. El problema con los detectores perceptuales es que la diferencia entre dos representaciones empleadas como detectores perceptuales en dos situaciones diferentes depende de una diferencia en la apariencia de un individuo (o de dos individuos de una misma clase) que puede *persistir* luego de que conozcamos la identidad relevante y que por lo tanto, *la diferencia de poder causal de esos detectores se pierda*. De este modo, considero que la teoría de proxytipos no proporciona una explicación adecuada de los casos Frege.

La propuesta de Prinz tiene este problema porque los conceptos pueden ser empleados como detectores perceptuales y, cuando lo son, su estructura semántica está determinada por la apariencia del objeto percibido ¿Pero qué sucede con la explicación de los casos Frege a partir de estructuras internas semánticamente complejas *que no sean perceptuales*? Fodor, justamente, considera que los conceptos no son representaciones perceptuales, aunque su activación dependa de la de estas representaciones. Vimos que, aún para los conceptos complejos, no podemos ofrecer una explicación sintáctica de los casos Frege. Sin embargo, Fodor podría dar una explicación semántica de los casos Frege para el caso de conceptos complejos, entendidos como estructuras no perceptuales. Esta propuesta parece poder explicar el fenómeno de la variación cognitiva. Veamos cómo. El concepto CLARK podría ser una estructura de Louis que, antes de que Louis conozca la identidad del superhéroe, tiene una estructura interna compuesta por conceptos como PERIODISTA DEL DAILY PLANET, CRIADO EN KANSAS, etc. Cuando CLARK es activada a partir de la percepción, esta activación depende de *otra* estructura, un detector perceptual cuya estructura interna es ANTEOJOS MARRONES, PELO HACIA ATRÁS, etc., es decir representaciones de las propiedades de apariencia de Clark. Por otro lado, antes de conocer la identidad, el concepto SUPERMAN de Louis es un concepto cuya estructura interna incluye

NACIDO EN KRIPTON, VUELA etc. La activación de SUPERMAN a partir de la percepción depende de la activación de una estructura perceptual cuya estructura interna es JOPO, CAPA ROJA, TRAJE AZUL, etc. Ahora bien, una vez que Louis conoce la identidad del superhéroe, aunque CLARK y SUPERMAN sigan siendo activados por medio de representaciones perceptuales diferentes, con una estructura interna diferente, CLARK y SUPERMAN mismos puede pasar a tener la misma estructura interna. Ambos conceptos pueden pasar a tener una estructura compuesta por CRIADO EN KANSAS, NACIDO EN KRIPTON, PERIODISTA DEL DAILY PLANET, VUELA, etc. El hecho de que los dos conceptos pasen a tener la misma estructura semántica interna explica que sus poderes causales pasen a ser los mismos. Aquí tendremos una explicación informacional de los casos Frege basada en la estructura interna de los conceptos que permite la explicación de la variación fregeana.

Considero que esta explicación no es adecuada. Como vimos, lo que determina el poder causal de una representación no es el contenido informacional mismo, sino ciertas propiedades neurológicas o computacionales intrínsecas de las representaciones, es decir, la forma sintáctica. Aunque, como vimos, no admito que la forma sintáctica nos sirva para explicar los casos Frege, dada la teoría computacional de la mente, es cierto que las propiedades de las representaciones a las que responde la mente son las propiedades sintácticas. Además, el que una estructura tenga cierto contenido informacional no determina que tenga cierta forma sintáctica ni viceversa. Esto es así porque, como argumenta Fodor (1994), las propiedades sintácticas son propiedades intrínsecas de las representaciones, mientras que las propiedades informacionales son propiedades extrínsecas o relacionales de las mismas y la posesión de una propiedad intrínseca no determina necesariamente (con necesidad metafísica o conceptual) la posesión de una propiedad extrínseca, ni viceversa (1994, cap. 1). Pero entonces, dos conceptos con una misma estructura semántica interna pueden tener una estructura sintáctica diferente. Si esto ocurre, estos conceptos tendrán poderes causales diferentes y por lo tanto tendremos que decir que son conceptos diferentes. Esto quiere decir que podemos tener casos Frege para conceptos que, además de tener la misma referencia, tienen la misma estructura semántica interna. Por lo tanto, necesitamos un criterio más restrictivo para individuar a los conceptos. De esto podemos concluir que no hay una explicación adecuada de los casos Frege en base a la estructura interna de los conceptos, sea sintáctica o semántica, perceptual o no perceptual.

Es interesante notar que, por el contrario, cualquier propuesta de rol funcional puede dar cuenta del fenómeno de la variación fregeana sin tener el problema recién mencionado de las estructuras internas semánticas. Una vez que Louis conoce la identidad entre Superman y Clark asociará la misma

información a ambos conceptos. Ambos conceptos pasarán a tener el mismo rol funcional. Louis sabrá que Clark vuela y que Superman fue criado en Kansas. CLARK estará causalmente relacionado con VUELA y SUPERMAN estará causalmente relacionado con CRIADO EN KANSAS. Lo mismo vale para cualquier otra información asociada a alguno de estos conceptos. De este modo, luego que Louis conoce la identidad, una propuesta de rol funcional puede explicar por qué los conceptos pasan a ser idénticos. CLARK y SUPERMAN han pasado a ser cognitivamente equivalentes porque tienen el mismo rol funcional. Así, esta propuesta no tiene el problema que tienen las explicaciones a partir de la estructura semántica interna perceptual.

Pero las propuestas de rol funcional tampoco tienen el problema de las estructuras internas semánticas no perceptuales. Si el contenido es identificado con el rol funcional, no tendremos casos en los que dos conceptos tengan diferentes poderes causales pero nuestro criterio de individuación determine que son instancias del mismo tipo. La razón es simplemente que el rol funcional *no es otra cosa que la especificación de esos poderes causales*. Ahora bien, he mostrado en el capítulo anterior que aunque los enfoques funcionalistas no adolecen, como piensa Fodor, del problema de la composicionalidad, no pueden cumplir con el requisito de publicidad. Sin embargo, argumenté en el capítulo anterior que si cruzamos un enfoque funcional con uno informacional podemos resolver este problema. Repasemos un poco los puntos centrales de mi propuesta para considerar si, al igual que las propuestas de rol funcional, puede evitar los problemas considerados en este capítulo.

Según el enfoque que propuse, hay dos elementos que constituyen el contenido cognitivo de un concepto dado C : (i) La descripción del rol funcional de C , que nos permite identificar un conjunto de representaciones $\{Q\}$ relacionadas inferencialmente con C , y (ii) los contenidos informacionales de las representaciones en $\{Q\}$. Dije que el contenido cognitivo de C no es el contenido informacional de C mismo, sino que es *el contenido informacional de las representaciones en $\{Q\}$* , el contenido informacional de las representaciones con las que C se relaciona inferencialmente. Pero esto quiere decir que si bien el contenido cognitivo de C no se identifica con su propio contenido informacional, sí se identifica con ciertos contenidos de este tipo. En este sentido la propuesta que ofrezco es una teoría causal o informacional del contenido cognitivo. Como dije antes, es una propuesta de contenido informacional *en* el rol funcional.

Como señalé en el capítulo pasado, para explicar los casos Frege empleo una teoría causal compleja. Por un lado, el contenido cognitivo de un concepto C está determinado por los contenidos

informacionales de algunas de las representaciones que aparecen en su descripción funcional. Por otro lado, la referencia de C sí será identificada con su propio contenido informacional. Por ejemplo, tanto SUPERMAN como CLARK se relacionan causalmente con el mismo individuo y por lo tanto tienen la misma referencia. Sin embargo, están relacionados inferencialmente con representaciones que tienen diferentes contenidos informacionales. SUPERMAN (pero no CLARK) está relacionado inferencialmente con una representación que se relaciona nomológicamente con la propiedad de volar y CLARK (pero no SUPERMAN) está relacionado inferencialmente con una representación que se relaciona nomológicamente con la propiedad de ser criado en Kansas. Esto hace que SUPERMAN y CLARK tengan contenidos cognitivos diferentes.

Repasemos cómo esta propuesta puede resolver el problema de la publicidad de los conceptos. Dije que el contenido cognitivo de un concepto C estará determinado sólo por los contenidos informacionales de las representaciones que aparecen en las inferencias por default en las que figura C . Vimos que apelar a inferencias por default no bastaba para que una propuesta de rol funcional resuelva el problema de la publicidad porque los roles así reducidos aún incluyen un conjunto de inferencias demasiado amplio. El problema era que la lista reducida de inferencias en las que entra un concepto C incluye conceptos que deben ser a su vez identificados por sus roles funcionales restringidos. A su vez, estos nuevos roles funcionales introducirán nuevos conceptos que deben ser definidos funcionalmente y así sucesivamente. En la propuesta que ofrecí, el contenido de C es determinado también por los contenidos de un conjunto de representaciones $\{Q\}$ que aparece en la descripción funcional de C . Sin embargo, los contenidos de las representaciones en $\{Q\}$ no son, a su vez, definidos funcionalmente, sino por medio de una teoría causal o informacional. No se apela a su relación con otras representaciones sino con un estímulo distal. Esto hace que la lista relevante de inferencias no siga extendiéndose a pesar de haber reducido el rol funcional original (la lista de inferencias en las que aparece C). De este modo, es más plausible que la propuesta de contenido informacional en rol funcional pueda satisfacer el requisito de publicidad.

Ahora bien, considero que mi propuesta, a pesar de no ser una propuesta de rol funcional, puede explicar el fenómeno de la variación fregeana. Esto es simplemente porque, como vimos la variación fregeana es determinada por la variación de las relaciones inferenciales de las representaciones. Cuando Louis conoce la identidad de Superman, sus conceptos CLARK y SUPERMAN pasan a tener las mismas relaciones inferenciales. Pero según mi propuesta, las relaciones inferenciales también determinan, aunque indirectamente, el contenido. Aunque el contenido no es idéntico a un conjunto de

relaciones inferenciales, es idéntico a la información contenida en las representaciones que aparecen en las relaciones inferenciales de un concepto. Por lo tanto, si dos conceptos comparten relaciones inferenciales, tendrán el mismo contenido cognitivo según mi propuesta. Así la propuesta de información en el rol funcional no adolece el problema de la explicación de los casos Frege en base la estructura interna semántica y perceptual de los conceptos.

En relación al problema de la explicación de los casos Frege por la estructura semántica interna no perceptual, considero que la propuesta de información en el rol funcional tampoco lo adolece. Creo que no se dará el caso que conceptos con el mismo contenido cognitivo tengan diferentes poderes causales. Esto es, nuevamente porque esos poderes causales son principalmente las relaciones inferenciales en las que entra un concepto y si bien en mi propuesta estas inferencias no se identifican con el contenido, lo determinan. Si una representación *P* tiene relaciones inferenciales que otra *Q* no, esto quiere decir que *P* se vincula causalmente con al menos una representación *R* con la que la *Q* no. Pero entonces, el contenido informacional de *R* será parte del contenido cognitivo de *P* pero no de *Q*.

Así, podemos concluir, de acuerdo con Dretske, Fodor y Prinz en que podemos explicar los casos Frege a partir de la estructura de los conceptos correferenciales involucrados. Pero a diferencia de ellos considero que esta estructura no es interna, sino lo que podemos llamar una estructura “externa”, el conjunto de contenidos informacionales en el rol funcional.

En estos dos primeros capítulos he intentado motivar una explicación informacional del contenido conceptual, tanto intencional como cognitivo y defenderla en contra de las alternativas informacionales más relevantes. En los tres capítulos que siguen intentaré extender el enfoque informacional más allá de las representaciones conceptuales. En primer lugar (capítulo 4), trataré de determinar el vínculo entre la información y las representaciones perceptuales. En segundo lugar, ofreceré una propuesta informacional de la noción de conocimiento (capítulo 6). Pero además, determinaré vínculos entre estos tipos de representaciones informacionalmente caracterizados. En el capítulo 5 mostraré diferentes relaciones conceptuales entre las nociones de representación conceptual y representación perceptual. En el capítulo 6 vincularé la noción de conocimiento con las nociones de representación perceptual y conceptual defendidas.

Capítulo 4:

Representaciones perceptuales

1. Introducción

En este capítulo intentaré elucidar la noción de representación perceptual y determinar la relación entre esta noción y la de información. La idea central es ofrecer una caracterización no semántica de las representaciones perceptuales, en base a la noción de sistema perceptual, pero mostrar que esta noción tiene ciertas implicaciones para la semántica de estas representaciones. La idea de caracterizar a lo perceptual a partir de los sentidos está inspirada en una estrategia semejante llevada adelante por Prinz (2002). Más aún, mi propuesta fue desarrollada en un intento de profundización de este enfoque. Llamaré a esta propuesta respecto de las representaciones perceptuales, la caracterización de tipos de representaciones según la facultad a la que pertenecen, “el enfoque de las facultades”. En este capítulo, la argumentación a favor de mi propuesta partirá de una reflexión crítica sobre el enfoque de las facultades de Prinz. Hay dos diferencias cruciales entre el abordaje de Prinz y el aquí propuesto. En primer lugar, argumentaré en contra de la caracterización que Prinz ofrece de los sistemas perceptuales. Prinz considera principalmente dos propuestas de caracterización de los sistemas perceptuales, la propuesta modular y la propuesta de la receptividad. Argumenta que sólo funciona una combinación de ambas. Por el contrario argumentaré que ninguno de los elementos de la propuesta modular nos sirve para distinguir adecuadamente a los sentidos de otras facultades y que nos basta con el elemento que propone el enfoque de la receptividad. En segundo lugar, criticaré el modo en que Prinz relaciona las nociones de representación y sistema perceptual y ofreceré una caracterización alternativa de esta relación. En tercer lugar, argumentaré que la noción de representación perceptual implica la de información pero no viceversa. Este tercer punto es importante ya que muestra que, por principio, no hay una caracterización semántica de las representaciones perceptuales. El contenido que las representaciones perceptuales tienen necesariamente, un contenido informacional, no es un

contenido característico o exclusivo de ellas, sino un tipo de contenido que pueden poseer representaciones no perceptuales. Prinz opone su enfoque de las facultades a los enfoques semánticos que se han propuesto tradicionalmente para caracterizar a las representaciones perceptuales y argumenta en contra de cada uno en particular. En este capítulo, por el contrario, en base a la relación entre las nociones de representación perceptual e información, ofrezco un argumento general en contra de estas caracterizaciones.

En la segunda sección presentaré el enfoque de Prinz sobre los sentidos y las representaciones perceptuales. En la sección tercera criticaré la noción de Prinz de sistema perceptual y defenderé el llamado “enfoque de la receptividad”. Por último, en la cuarta sección, defenderé una caracterización particular de la relación entre los sistemas perceptuales y las representaciones perceptuales y determinaré la relación entre la noción de representación perceptual y la de información.

2. Sistemas perceptuales

Como afirmé antes, disiento con Prinz tanto en su caracterización de los sistemas perceptuales como en el modo en que conecta estos sistemas con las representaciones perceptuales. En esta sección desarrollaré brevemente la estrategia de Prinz y expondré con algo de detalle su caracterización de los sistemas perceptuales.

Prinz (2002) considera que podemos caracterizar a las representaciones perceptuales como representaciones que pertenecen a sistemas perceptuales. Para distinguir a las representaciones perceptuales de otro tipo de representaciones, si las hubiere, debemos distinguir a los sentidos de otras facultades cognitivas. Antes de ofrecer su caracterización de estos sistemas, Prinz discute varias propuestas alternativas de caracterización de las representaciones perceptuales. Discute centralmente diversos criterios sintácticos y semánticos. La motivación de su propuesta se apoya tanto en la intuición de su criterio en base a sistemas perceptuales como en el rechazo de estos criterios alternativos. Por el contrario, en la tercera sección presentaré un argumento directo para este enfoque, independiente de la discusión de estos criterios alternativos.

La idea de Prinz (2002) es, entonces, que una representación perceptual no es otra cosa que una representación de nuestros sentidos. Prinz considera dos propuestas principales para caracterizar a los

sentidos. Son lo que llama la “propuesta modular” y “la propuesta de la receptividad”. La propuesta modular afirma que los sistemas sensoriales, a diferencia de otros sistemas, son modulares (Fodor 1983). La propuesta de la receptividad caracteriza a los sentidos por medio de lo que considera su función primaria, recibir inputs del entorno extramental, incluyendo el propio cuerpo. Prinz cree que las dos propuestas tienen problemas pero que una combinación de ellas puede funcionar.

¿Qué quiere decir que los sistemas perceptuales son modulares? Hay cuatro características asociadas con los sistemas modulares. Son rápidos, de dominio específico, asociados con arquitecturas neuronales específicas e informacionalmente encapsulados. Prinz rechaza la noción de módulo perceptual porque niega que los sistemas perceptuales estén informacionalmente encapsulados. Pero toma de esta concepción la idea de que los sistemas perceptuales son de dominio específico (en diferentes sentidos que especificaré más adelante) y que tienen una arquitectura neuronal específica.

Decir que los sistemas perceptuales son informacionalmente encapsulados significa que el procesamiento de los sistemas perceptuales no puede ser influenciado por la información contenida en otros sistemas. Se ha empleado evidencia empírica para argumentar que los sistemas perceptuales no son informacionalmente encapsulados y por lo tanto, tampoco modulares (por ejemplo, Churchland 1988). Fenómenos bien conocidos como el de la restauración de fonemas (en el cual escuchamos sonidos que no son efectivamente articulados por un hablante) y el de la interpretación de las imágenes fragmentadas (el cuál a veces depende del conocimiento de los items representados) proporciona evidencia a favor de que las creencias, las expectativas y las interpretaciones de nivel superior pueden afectar lo que percibimos. Evidencia adicional en contra del encapsulamiento viene del hecho de que hay muchas vías neuronales eferentes de regiones superiores del cerebro a las regiones perceptuales.

En relación con la propuesta de la receptividad, Prinz considera primero una versión muy implausible (aunque tradicional) de esta propuesta: La idea kantiana de que los sentidos son receptivos, que reciben pasivamente estimulaciones del entorno, mientras que el intelecto es espontáneo, generando activamente nuevas representaciones bajo el control del organismo. Esta propuesta enfrenta diversos problemas. El más obvio surge de lo que he dicho recién sobre encapsulamiento informacional, Es ampliamente sostenido que transformamos, reconstruimos e interpretamos señales desde estadios de procesamiento muy tempranos, reduciendo ruido o información irrelevante, completando información faltante utilizando conocimiento previo y expectativas. Así, nuestros sentidos no son meramente pasivos si esto significa que no interpretan, transforman o reconstruyen el input al

que responden.

Prinz considera que a pesar de sus dificultades, ambas propuestas están en lo correcto en relación con ciertos puntos. Toma de la propuesta de la receptividad la idea de que los sentidos tienen la función de responder a inputs. Toma de la propuesta modular la idea de que los sentidos sirven a funciones que son de dominio específico. A partir de estas ideas propone la tesis de que los sentidos son sistemas que responden a clases particulares de inputs. De un modo más específico, afirma que los sentidos son sistemas dedicados a un input específico. Veamos ahora cómo caracteriza cada una de las características constitutivas de su propuesta.

Al decir que los sentidos son *sistemas* Prinz quiere decir tres cosas diferentes. Las tres son características asociadas con la noción de modularidad. Las primeras dos están relacionadas con el carácter de dominio específico de los módulos. Afirma que cada sentido tiene su propio conjunto de operaciones y que cada sentido tiene su propio conjunto de representaciones. La tercer característica está relacionada con la arquitectura neuronal: Las operaciones y representaciones tienen vías neuronales específicas.

Al decir que los sentidos son sistemas *de input*, Prinz sólo quiere decir que son sistemas que reciben inputs de afuera del cerebro. Esta es una versión mucho más plausible de la propuesta de la receptividad. Los inputs pueden venir del entorno externo (como en la audición, la vista y el sistema de feromonas) o de adentro del cuerpo (como en la propiocepción, interocepción, el hambre y la sed). Prinz afirma que algunos componentes de un sistema perceptual yacen lejos de los transductores que reciben la señal input, pero para contar como un sentido, debe haber inputs en algún lugar y los componentes deben contribuir al procesamiento de estos inputs.

Finalmente, decir que los sentidos son sistemas *dedicados* quiere decir, para Prinz, dos cosas. La primera es sólo una especificación de la condición previa. No es sólo el caso de que podemos distinguir a un sentido de algo que no lo es determinado si hay un input externo al cerebro sino que también podemos distinguir un sentido de otro por medio del tipo de input al que responde. Cada sentido responde a un tipo de input particular. Se sabe que cada sentido es especialmente apto para captar diferentes tipos de magnitudes físicas. Por ejemplo, la visión responde a la longitud de onda de la luz, la audición responde a la frecuencia de movimiento molecular y el olfato responde a formas moleculares. La segunda característica que Prinz liga a la noción de dedicación es nuevamente la relacionada con la especificidad de dominio de los módulos perceptuales. La asunción de que los

diferentes sentidos usan diferentes tipos de representación.

En la sección que sigue presentaré una crítica a los aspectos de la propuesta de Prinz que surgen de la propuesta modular, defenderé una propuesta de la receptividad y ofreceré una crítica al modo de Prinz de caracterizar a las representaciones perceptuales a partir de los sentidos.

3. Una defensa de la propuesta de la receptividad

Argumentaré ahora a favor de la principal intuición detrás de las propuestas de la receptividad, la idea de que los sistemas perceptuales son plenamente caracterizados por la función a la que sirven. Para ello quiero especificar dos puntos que considero claves para una elucidación adecuada de la noción de sistema perceptual.

En primer lugar, creo que es crucial que el rasgo que determinemos como constitutivo de los sistemas perceptuales sea relevante para distinguir un sentido de algo que no es un sentido, de otras facultades cognitivas. Me interesa determinar *qué rasgos hacen que un sistema sea específicamente perceptual y no de otro tipo*.

En segundo lugar considero que una caracterización de la noción de sistema perceptual debe ser lógicamente compatible con la tesis empirista, tal como la plantea Prinz, es decir, la tesis de que todos los conceptos son representaciones perceptuales o copias de representaciones perceptuales. Creo que esto es recomendable sea cual sea la posición que tengamos frente a esta tesis. Este punto es obvio para quien sostiene que la tesis empirista es verdadera y para quien sostiene que es una tesis empírica, una tesis que no puede ser falsa sólo por los conceptos que involucra. Pero además, considero que quien sostiene que la tesis es falsa, está en una situación dialécticamente ventajosa si puede aceptar una noción de sistema y de representación perceptual que no implique esa falsedad. Éste último es mi caso. En el capítulo que sigue argumentaré en contra de la tesis empirista. Por lo tanto, considero que si puedo ofrecer una noción de sistema perceptual y de representación perceptual que sea lógicamente compatible con ella, el argumento en contra del empirismo tiene más fuerza. En esta sección argumentaré que los criterios que Prinz toma de las propuestas modulares, son conceptualmente incompatibles con el empirismo de conceptos.

Empecemos considerando los primeros dos rasgos que Prinz toma de las propuestas modulares, las representaciones y operaciones propias de los sistemas perceptuales y evaluemos la tesis de que estas propiedades caracterizan a los sistemas perceptuales. En relación con las representaciones, uno puede adoptar una caracterización semántica, sintáctica, funcional o una combinación de ellas. Como hemos visto, las propiedades sintácticas de las representaciones son típicamente caracterizadas como propiedades físicas intrínsecas de las representaciones (Fodor 2008). Por su parte, parece razonable caracterizar a las operaciones sobre representaciones como relaciones causales sistemáticas entre representaciones, es decir, como los roles funcionales de las representaciones. Así, la idea de caracterizar a los sentidos por su peculiar conjunto de representaciones y operaciones puede reducirse a la idea de caracterizarlos por medio de las propiedades semánticas, físicas y/o funcionales de las representaciones que manipulan. Podemos llamar a esta tesis “el enfoque representacional de los sentidos”.

El enfoque representacional afirma que un sistema sólo cuenta como perceptual si tiene un conjunto distintivo de representaciones (en el sentido semántico, sintáctico y/o funcional). Pero “distintivo” puede querer decir dos cosas: (i) diferente de las representaciones de otros sistemas perceptuales y (ii) diferente de las representaciones en sistemas no perceptuales. Ahora bien, si queremos un enfoque que nos permita distinguir un sistema perceptual de uno no perceptual, lo que nos interesará es “distintivo” en el sentido de (ii). Sólo si las representaciones en sistemas perceptuales son diferentes de las representaciones en sistemas no perceptuales podremos distinguir estos sistemas por medio de sus representaciones. Ahora bien, la tesis del empirismo de conceptos es incompatible con este criterio. El empirismo de conceptos afirma, según Prinz (2002), que todos los conceptos son copias o combinaciones de copias de representaciones perceptuales (p. 108). Aunque consideremos con un poco más de detalle la noción de copia más adelante, podemos adelantar que una copia es o bien una reproducción de una representación perceptual en otro sistema o bien un registro, en otro sistema, que permite reactivar la representación perceptual en el sistema perceptual en ausencia de estímulos externos. Interpretando de cualquiera de estas dos formas la noción de copia, es claro que los sistemas no perceptuales emplean, según esta tesis, las mismas representaciones que los sistemas perceptuales. Según la primera interpretación son instancias diferentes del mismo tipo (una representación perceptual es instanciada en otro sistema no perceptual) y según la segunda interpretación se trata de la misma instancia (una representación perceptual es reactivada, para ser empleada por otro sistema). Como vimos, considero que una noción de sistema y de representación perceptual compatible con el

empirismo es preferible incluso para el que quiere argumentar, como lo haré en el próximo capítulo, en contra de él. Por lo tanto las representaciones de los sistemas perceptuales no pueden ser empleadas para distinguirlos de otros sistemas.

¿Qué sucede con la tercer característica que Prinz toma de la propuesta modular: las vías neuronales? Prinz considera que distinguir vías neuronales separadas es crucial ya que la propuesta alternativa, de acuerdo con la cual podemos individuar sentidos por los órganos sensoriales externos, tiene diversos problemas. Un problema es que podemos imaginar criaturas que tengan dos órganos que sirvan a una misma modalidad. Un ejemplo es la interocepción. Los órganos internos del cuerpo son diferentes, pero proporcionan inputs al sistema sensorial que monitorea sentimientos internos. Otros problema es que el mismo órgano puede servir a diferentes sentidos. Por ejemplo, la piel sirve en la detección de calor y de presión, que pueden ser considerados diferentes sentidos.

Prinz considera una propuesta alternativa. Podemos identificar a los sentidos por sus diferentes tipos de receptores. Los diferentes sentidos tienen diferentes tipos de receptores y estos desempeñan un papel importante en la individuación de los sentidos. Sin embargo, Prinz considera que sería un error detenernos en los receptores. Dice que podemos imaginar criaturas que tengan muchos tipos de receptores afuera del cerebro pero un sólo sistema de procesamiento en el cerebro. (Una versión extrema de la propuesta de la equipotencialidad de Lashley 1950). Los verdaderos sistemas sensoriales pueden ser distinguidos internamente. Prinz afirma que los sentidos deben ser identificados por distintas colecciones de poblaciones de neuronas cooperativas.

Aún si concedemos el argumento de Prinz, hay dos puntos importantes en contra de emplear este rasgo de los sistemas perceptuales. Supongamos que Prinz está en lo cierto y los diferentes sistemas perceptuales deben distinguirse entre sí, en parte, por tener diferentes vías neuronales. Aún si esto es así, en primer lugar, esto no implica que podamos distinguir un sistema perceptual *de uno no perceptual* por su vía neuronal. Prinz no ofrece argumentos a favor de esta tesis sino sólo a favor de que los sistemas perceptuales son distinguidos *entre sí* por sus vías neuronales diferentes. Pero este argumento sólo nos proporciona un modo de distinguir entre diferentes sentidos, no entre un sentido y algo que no lo es.

En segundo lugar, y más importante, aún si hubiese un argumento o evidencia empírica a favor de la tesis de que los sistemas perceptuales tienen vías neuronales diferentes de las que tienen los sistemas no perceptuales, este rasgo no nos dice *por qué una vía neuronal es perceptual y otra es no*

perceptual y por lo tanto, tampoco puede ser usado como criterio para distinguir a los sentidos de otras facultades. Notar que hay dos vías neuronales diferentes nos permite, si aceptamos la tesis de que todo sistema tiene una vía neuronal propia, determinar que hay dos sistemas diferentes. Pero no podemos determinar, en base a esta distinción de vías, si alguna de ellas es perceptual y cuál. Para tener un criterio adecuado basado en las vías neuronales deberíamos ofrecer una propiedad específica, propia de las vías neuronales perceptuales. Por ejemplo, deberíamos mostrar que hay alguna diferencia en la tasa o tiempo de disparo neuronal entre las vías neuronales perceptuales y las no perceptuales.

Ahora bien, creo que, nuevamente, un criterio de este tipo es conceptualmente incompatible con la tesis empirista. Las diferencias entre las vías neuronales que estamos considerando tienen que ser diferencias que tengan alguna relevancia cognitiva. Queremos saber qué hace a una vía neuronal perceptual y qué la hace no perceptual, sino perteneciente a otro sistema cognitivo. Ahora bien, los sistemas cognitivos son sistemas que manipulan representaciones. Las propiedades neuronales que nos interesarán son entonces las propiedades causalmente relevantes para la manipulación de representaciones, es decir, serán propiedades neuronales de las representaciones. Son las propiedades neuronales de las representaciones a las que estos sistemas responden.

Como vimos, la tesis empirista afirma que los sistemas no perceptuales operan sobre copias de representaciones perceptuales. Considero que según cualquiera de las dos nociones de copia que considera Prinz, esta tesis implica que no habrá diferencias neuronales relevantes entre los sistemas perceptuales y los no perceptuales. Una primera propuesta sobre la noción de copia es que copiar es reproducir en un área no perceptual del cerebro, *un patrón de actividad neuronal* de algún área perceptual del cerebro (p. 108). Según esta primera alternativa no habría diferencias neuronales relevantes entre los sistemas perceptuales y los no perceptuales pues lo reproducido son las propiedades neuronales mismas de la representación. Según una segunda propuesta, una copia podría ser un registro que deja una representación perceptual en un sistema no perceptual y que permite que la representación perceptual sea regenerada *en el sistema perceptual* en ocasiones subsiguientes, en ausencia de un estímulo externo. Como señala Prinz, los registros no son copias, sino instrucciones para producir copias. Ahora bien, esta alternativa sí permite que haya una diferencia a nivel neuronal entre una vía perceptual y una no perceptual. Estos registros en los sistemas no perceptuales podrían tener propiedades neuronales diferentes de las representaciones que regeneran en los sistemas perceptuales. Sin embargo, como señalé previamente, los sistemas no perceptuales que estamos considerando son sistemas cognitivos, sistemas cuyas operaciones deben ser caracterizadas como operaciones sobre

representaciones. Esto quiere decir que los estados neuronales relevantes para caracterizar las operaciones de un sistema no perceptual no serán los registros mismos, sino las representaciones que reproducen en los sistemas perceptuales. Por lo tanto, en lo que respecta a los aspectos cognitivamente relevantes, no debería haber diferencias neuronales relevantes tampoco según esta alternativa. Por estos motivos considero que si afirmamos que los sistemas perceptuales y los no perceptuales deben ser distinguidos por sus propiedades neuronales, la tesis empirista se vuelve falsa por la noción misma de sistema perceptual. El criterio para distinguir entre estos sistemas cognitivos hace conceptualmente imposible que los sistemas *no perceptuales* empleen copias de las representaciones en los sistemas perceptuales.

Podemos concluir entonces que no podemos apelar a ninguna de las características que Prinz toma de las propuestas modulares (representaciones, operaciones y vías neuronales) para distinguir a los sentidos de otras facultades ¿Qué sucede con el elemento de las propuestas de la receptividad, la función de responder a un input específico? En principio, parece que este criterio, a diferencia de los otros, no entra en conflicto con una perspectiva empirista. La tesis empirista afirma que las copias de representaciones perceptuales son empleadas por los sistemas no perceptuales, pero no que las empleen *para cumplir con las mismas funciones*, para ejecutar las mismas tareas. Por lo tanto, considero que este criterio, al no ser conceptualmente incompatible con la tesis empirista, es más aceptable que los que provienen de la propuesta modular. Concluyo entonces que el criterio propuesto por el enfoque de la receptividad permite distinguir adecuadamente a los sistemas perceptuales de los no perceptuales. Estos sistemas pueden distinguirse según su función específica. La propiedad de los sistemas perceptuales relevante para distinguirlos de otros sistemas cognitivos es su función de responder a un input específico externo al sistema nervioso.

4. Representaciones perceptuales y contenido perceptual

Como mostré al comienzo de este capítulo, el interés de Prinz por los sistemas perceptuales es emplear la noción de sistema perceptual para elucidar la noción de representación perceptual. Las representaciones perceptuales no son otra cosa, según Prinz, que representaciones que pertenecen a sistemas perceptuales. Prinz considera que esta es la alternativa más viable para caracterizar a las representaciones perceptuales porque cree que las propuestas sintácticas y semánticas tienen problemas

graves. En esta sección, tras varias consideraciones críticas, voy a proponer una versión del criterio de Prinz para caracterizar a las representaciones perceptuales y estableceré, en base a este criterio, la relación entre la noción de representación perceptual y la de información.

Empecemos primero por las representaciones perceptuales. Si las representaciones perceptuales son las representaciones que pertenecen un sistema perceptual, la pregunta que surge naturalmente entonces, es qué quiere decir “pertenencia”. Como vimos, Prinz considera que decir que cada uno de los sentidos tiene su propio conjunto de representaciones quiere decir, al menos en parte, que las representaciones que pertenecen a los sentidos tienen propiedades semánticas, sintácticas y/o funcionales específicas. Sin embargo, hemos visto que, dada la tesis empirista, las representaciones de los sistemas no perceptuales tendrán las mismas propiedades. Por lo tanto, si pensamos, como argumenté, que es preferible una noción de representación perceptual lógicamente compatible con el empirismo, no podemos valernos de este criterio de pertenencia. Además, aún si nouviésemos este problema, si usamos a las representaciones perceptuales, como quiere Prinz, para caracterizar a los sentidos, entonces resultaría circular usar a los sentidos para caracterizar a las representaciones perceptuales. No podemos caracterizar a las representaciones perceptuales como el conjunto de representaciones que pertenecen a los sentidos y caracterizar al mismo tiempo a los sentidos como sistemas que poseen ese conjunto peculiar de representaciones. No podemos tener a la vez una teoría representacional de los sentidos y una teoría “sensorial” de las representaciones perceptuales.

Ahora bien, como he mostrado, no necesitamos caracterizar a los sentidos en función de sus representaciones. Podemos caracterizarlos plenamente en virtud de su función peculiar. Tal vez podemos usar esta caracterización para elucidar el sentido adecuado de “pertenencia a un sentido”. Podemos decir que las representaciones perceptuales son las que contribuyen a la función de los sistemas perceptuales. Podemos decir que las representaciones que pertenecen a un sistema perceptual son las representaciones que tienen la función de responder a un input específico externo al sistema nervioso. Es importante notar que, si bien he ofrecido en la sección anterior argumentos independientes en contra de caracterizar a los sentidos por medio de sus representaciones y a favor de caracterizarlos sólo en base a su función, habla a favor de esta última propuesta el hecho de que, cómo vimos en el párrafo anterior, sólo a partir de ella es posible usar a los sentidos para elucidar la noción de representación perceptual.

Ahora bien, considero que este criterio, si bien encaminado, es aún insuficiente. Como vimos, la

tesis empirista implica que las representaciones perceptuales sirven tanto a la función de sistemas no perceptuales como a la de los sistemas perceptuales. Esto quiere decir que no podemos llamar a ciertas representaciones “perceptuales” por el hecho de que contribuyen a la función de los sistemas perceptuales y decir que los sistemas no perceptuales emplean representaciones perceptuales. Pues con el mismo criterio podemos decir que las representaciones son “no perceptuales” porque contribuyen a la función de los sistemas no perceptuales y afirmar que los sistemas perceptuales emplean representaciones no perceptuales. Este criterio para caracterizar a las representaciones perceptuales no nos permite distinguir la tesis empirista de que las representaciones perceptuales son empleadas por todos los sistemas cognitivos, de la tesis racionalista de que los sentidos emplean las representaciones no perceptuales de las facultades cognitivas superiores.

Considero, sin embargo, que en Prinz mismo encontramos las herramientas conceptuales para decir por qué, a pesar de que todos los sistemas emplean las mismas representaciones, estas representaciones son perceptuales. El punto es simplemente que, si todos los sistemas cognitivos emplean las mismas representaciones, podemos decir que son perceptuales ssi (i) las representaciones en los sistemas no perceptuales son *copias* de representaciones previamente empleadas por los sistemas perceptuales y (ii) el empleo de estas representaciones por parte de los sistemas perceptuales no presupone el proceso de copiado. Esto quiere decir que el criterio de Prinz necesita no sólo de la caracterización de los sistemas perceptuales, sino también de la noción de copia. Las representaciones perceptuales son representaciones que sirven a la función de un sistema perceptual sin la mediación de un proceso de copiado. Por contraste, esto nos da un modo de caracterizar la noción de representación no perceptual. Una representación no perceptual es una representación que es empleada por un sistema no perceptual sin mediación de un proceso de copiado.

Estas caracterizaciones prescinden claramente de toda propiedad, semántica o sintáctica de las representaciones perceptuales. Prinz cree que esto es una ventaja ya que considera que las principales propuestas semánticas o sintácticas adolecen de graves problemas. En esta sección argumentaré, sin embargo, que esta noción de representación perceptual nos dice algo sobre la semántica de las representaciones perceptuales.

Hay un vínculo obvio entre la noción de información y la noción de representación perceptual propuesta. Si las representaciones perceptuales tienen la función de correlacionarse con inputs externos al sistema nervioso, portarán entonces información sobre los mismos. Esto es importante porque

implica que el hecho de que las representaciones perceptuales porten información no es un hecho contingente sino una verdad conceptual. Si la caracterización de la noción de representación perceptual propuesta es adecuada, entonces esta noción implica la de información. Ahora bien, a primera vista, parece que la implicación también va en la otra dirección. Vimos que la función de los sistemas perceptuales es lo que los distingue de otros sistemas. Pero si sólo las representaciones de los sistemas perceptuales se vinculan con inputs externos entonces sólo las representaciones perceptuales portan información sobre hechos externos al sistema nervioso. Es decir, no sólo ser una representación perceptual implica portar información sino que ser una estructura neuronal que porta información implica ser perceptual. Sin embargo, considero que esta última implicación no se da. Una estructura neuronal puede portar información sobre hechos externos al sistema nervioso sin ser una representación perceptual, sin pertenecer a un sistema perceptual. Veamos cómo.

Es claro que aunque dos representaciones que pertenezcan a sistemas perceptuales diferentes no responderán nunca a un mismo tipo de input proximal, sí pueden portar información sobre la presencia de un mismo estímulo distal. Tanto la representación *A* de un sistema auditivo que responde al patrón de movimiento molecular producido por un ladrido como la representación *V* de un sistema visual activada por un patrón de ondas de luz producido por el cuerpo de un perro, portarán información sobre la presencia de un perro. Ahora bien, podemos imaginar una estructura neuronal *N* que sea activada tanto por medio de la activación de *A* como por medio de la activación de *V*. La estructura *N* portará información sobre la presencia de un perro, pues sólo será activada en presencia de perros, pero al ser activada por diferentes tipos generales de input proximales no podemos decir que sea perceptual. Pues vimos que las representaciones perceptuales sólo son activadas en presencia de un tipo de input específico. Por lo tanto, podemos concluir que si bien la noción de representación perceptual implica la de información, la noción de información no implica la de representación perceptual.

Este argumento, además de elucidar la relación que hay entre las nociones de representación perceptual e información, constituye un modo peculiar de rechazar una caracterización semántica de las representaciones perceptuales. He mostrado que la noción misma de representación perceptual implica cierto tipo de contenido, un contenido informacional. Esto es interesante ya que, como he afirmado reiteradas veces, la noción de información parece la más apta para la naturalización del contenido. Sin embargo, como he mostrado que este contenido de las representaciones perceptuales puede ser poseído por representaciones no perceptuales, entonces no podemos emplearlo para caracterizar a lo perceptual. *Lo perceptual no puede ser caracterizado por el contenido que le es propio, ya que este contenido no*

es exclusivo de lo perceptual.

6. Conclusión

El vínculo estrecho entre las nociones de representación perceptual y de información que he elaborado en la última sección nos muestra que las representaciones conceptuales y las perceptuales tienen algo importante en común. En los capítulos 2 y 3 he desarrollado una propuesta informacional sobre las representaciones conceptuales. Si esta propuesta es adecuada, esto quiere decir que las representaciones perceptuales y las conceptuales son representaciones del mismo tipo, ambas poseen un contenido que podemos determinar en términos informacionales. Sin embargo, y esto es lo importante, he mostrado que informacional no implica perceptual (ni no perceptual). Además, hemos visto que no cualquier estructura que porta información es un concepto. En particular, argumenté que los proxytipos no lo son. Por lo tanto, informacional tampoco implica conceptual. Esto quiere decir que aún cuando los conceptos y perceptos tengan contenidos determinados informacionalmente, queda aún abierta la relación específica que hay entre estos tipos de representaciones. En el próximo capítulo me ocupo de examinar esta relación.

Capítulo 5:

La distinción y relación conceptual/perceptual y un anti-empirismo de conceptos

1. Introducción

En este capítulo considere algunos aspectos centrales de la relación entre lo conceptual y lo perceptual tal como lo he caracterizado en los capítulos anteriores. Específicamente, defenderé cuatro tesis sobre esta relación. Sostendré que (i) ser una representación conceptual no implica ser una representación perceptual, que (ii) ser una representación perceptual no implica ser una representación conceptual, que (iii) ser una representación no perceptual no implica ser una representación conceptual y finalmente, que (iv) ser una representación conceptual implica ser una representación no perceptual.

Las primeras dos tesis (i) y (ii) muestran que la tesis empirista no es verdadera en virtud de los conceptos que involucra. La tesis empirista, si fuese cierta, será a lo sumo una verdad empírica. La tesis (iii) nos muestra que el enfoque informacional de los conceptos y las representaciones perceptuales que he ofrecido deja abierta la posibilidad lógica de una representación que no sea perceptual ni conceptual. Por último, la tesis más fuerte (iv) implica que la tesis empirista es necesariamente falsa. En particular, argumentaré que una representación no puede ser a la vez perceptual y conceptual. Sostendré que si una representación perceptual adquiere contenido conceptual deja *ipso facto* de ser perceptual. Empecemos retomando algunos aspectos (relevantes para las tesis de este capítulo) de las propuestas desarrolladas sobre las representaciones conceptuales y perceptuales.

2. Representaciones conceptuales

Como vimos, ofrecí una propuesta informacional que integra roles funcionales y que es capaz de restringirlos adecuadamente. Según el enfoque que propuse hay dos elementos que constituyen el contenido de un concepto dado C : (i) La descripción del rol funcional de C , que nos permite identificar un conjunto de representaciones $\{Q\}$ relacionadas inferencialmente con C , y (ii) los contenidos informacionales de las representaciones en $\{Q\}$. Dije que el contenido cognitivo de C no es el contenido informacional de C mismo, sino que es *el contenido informacional de las representaciones en $\{Q\}$* , el contenido informacional de las representaciones con las que C se relaciona inferencialmente (que aparecen en sus inferencias por default). Pero esto quiere decir que si bien el contenido cognitivo de C no se identifica con su propio contenido informacional, sí se identifica con ciertos contenidos de este tipo. En este sentido la propuesta que ofrezco es una teoría informacional del contenido cognitivo.

Así, para explicar los casos Frege propuse una teoría informacional compleja. Por un lado, el contenido cognitivo de un concepto C está determinado por los contenidos informacionales de algunas de las representaciones que aparecen en su descripción funcional. Por otro lado, la referencia de C sí será identificada con su propio contenido informacional. Por ejemplo, tanto SUPERMAN como CLARK se relacionan informacionalmente con el mismo individuo y por lo tanto tienen la misma referencia. Sin embargo, están relacionados inferencialmente con representaciones que tienen diferentes contenidos informacionales. SUPERMAN (pero no CLARK) está relacionado inferencialmente con una representación que se relaciona informacionalmente con la propiedad de volar y CLARK (pero no SUPERMAN) está relacionado inferencialmente con una representación que se relaciona informacionalmente con la propiedad de ser criado en Kansas. Esto hace que SUPERMAN y CLARK tengan contenidos cognitivos diferentes.

En el capítulo 3 argumenté además en contra de propuestas informacionales alternativas. *Como las representaciones perceptuales tienen contenidos informacionales, es importante saber qué*

contenidos informacionales pueden y cuáles no pueden, ser conceptuales. En el capítulo 3 ofrecí argumentos en contra de propuestas informacionales que pretenden individuar a los conceptos en virtud de su estructura semántica *interna*. Argumenté, en particular, en contra de la idea de Prinz de que los casos Frege puedan ser explicados por la estructura semántica interna *de las representaciones perceptuales*.

Como expliqué en el capítulo 3, Prinz (2002) considera que el problema de explicar los casos Frege dentro de un enfoque informacional puede ser resuelto si entendemos a los conceptos como *proxytipos*. Los proxytipos son mecanismos que poseen una estructura interna compleja, que pertenecen a sistemas perceptuales y que pueden ser usados por esos sistemas para detectar propiedades. Prinz considera que la capacidad de detección de los proxytipos les da a los conceptos las propiedades informacionales que necesitan para dar cuenta del contenido intencional, mientras que la complejidad de la estructura permite explicar semánticamente los casos Frege, determinando para los proxytipos un contenido cognitivo.

Los proxytipos son estructuras almacenadas en la memoria de largo plazo que pueden ser empleadas como detectores cuando son activadas en la memoria de trabajo. En tanto detectores, son estructuras que pueden covariar nomológicamente con individuos o con miembros de una clase. Esta relación de covariancia con una propiedad dada permite, según la semántica informacional, que refieran a dicha propiedad. Pero además (a diferencia de lo que Prinz llama “indicadores”) estas estructuras son complejas: están compuestas de representaciones diferentes propiedades. Los proxytipos detectan alguna propiedad *P* que tienen como referencia detectando otro conjunto {*Q*} de propiedades superficiales o de apariencia que constituirán su contenido cognitivo (o, en terminología de Prinz, su contenido nominal). Esto quiere decir que proxytipos con la misma referencia pueden tener diferentes contenidos. Por ejemplo, CLARK puede ser un detector compuesto de ANTEOJOS, TRAJE MARRÓN, PEINADO HACIA ATRÁS, etc., mientras que SUPERMAN es un detector compuesto de CAPA ROJA, TRAJE AZUL, JOPO, etc. Son estructuras complejas que covarían con el mismo individuo (a quien por lo tanto refieren) covariando con diferentes conjuntos de propiedades suyas. De este modo, los proxytipos, siendo estructuras no atómicas, pueden ofrecer una explicación puramente informacional de los casos Frege.

Como vimos, La razón que motiva mi rechazo de la propuesta de Prinz es que considero que los proxytipos fallan en explicar lo que llamé “variación fregeana”, la situación en la que se pierde la diferencia entre dos conceptos típica de un caso Frege.

El problema que consideré que los proxytipos tienen para dar cuenta de la variación fregeana está directamente vinculada a su carácter perceptual, al hecho de que su activación *puede ser parte de la respuesta de un sistema perceptual a un input externo*. Si Louis tiene un proxytipo que usa para detectar perceptualmente a Clark Kent (una estructura, digamos, compuesta por ANTEOJOS, TRAJE MARRÓN, PELO HACIA ATRÁS, etc.) en un momento t_0 y otro diferente que usa para detectar perceptualmente a Superman (uno compuesto, por ejemplo, de CAPA, TRAJE AZUL, JOPO, etc.) en un momento t_1 , Prinz tendrá que atribuir a Louis dos conceptos con contenido cognitivo diferente en t_0 y t_1 . Ahora bien, es claro que aún después de que Louis sepa de la identidad del superhéroe, usará esos proxytipos diferentes para detectar perceptualmente al mismo individuo (pues dicho individuo seguirá teniendo una apariencia totalmente diferente en diferentes situaciones). Pero, como vimos, no podemos seguir atribuyendo a Louis en esas situaciones conceptos semántica o sintácticamente diferentes. Vimos que si Louis descubre la identidad del superhéroe, entonces los dos conceptos CLARK y SUPERMAN pasarán a ser cognitiva o sintácticamente equivalentes. Siempre podremos sustituir uno por el otro en contextos intensionales de atribución cuando Louis sea el sujeto. De este modo, el concepto atribuido en t_0 y t_1 no puede identificarse con las estructuras empleadas como detectores perceptuales en esos momentos. Como vimos, el problema con los detectores perceptuales es que la diferencia entre dos representaciones empleadas como detectores perceptuales en dos situaciones diferentes depende de una diferencia en la apariencia de un individuo (o de dos individuos de una misma clase) que puede *persistir* luego de que conozcamos la identidad relevante y que por lo tanto, *la diferencia de poder causal de esos detectores se pierda*. Por este motivo considero que la estructura semántica interna de las representaciones perceptuales no puede ser identificada con el contenido cognitivo y, por lo tanto, el contenido informacional de los proxytipos no puede ser identificado con el contenido conceptual. Si los proxytipos son conceptos, no lo son en virtud de su estructura semántica interna.

Como señalé en el capítulo 3, tampoco podemos ofrecer una explicación de los casos Frege a partir de estructuras internas semánticamente complejas *que no sean perceptuales*. Esto es porque lo que determina, en una situación fregeana, que dos conceptos sean diferentes, son crucialmente las diferencias causales de los mismos. Sin embargo, dado que la relación entre el contenido informacional de una estructura y su forma sintáctica es metafísicamente contingente, y que la forma sintáctica determina los poderes causales de un concepto, podemos tener un caso Frege para dos conceptos con la misma estructura semántica interna no perceptual. Podemos tener un caso en el que debamos distinguir conceptos con una misma estructura interna semántica. Por lo tanto, aún para el caso de las

representaciones no perceptuales, los conceptos no pueden ser individuados en base al contenido informacional de sus componentes internos. Si una representación no perceptual es un concepto, *esto no es determinado por su estructura semántica interna*.

Es importante recordar que, por el contrario, cualquier propuesta de rol funcional puede evitar estos problemas. Una vez que Louis conoce la identidad entre Superman y Clark asociará la misma información a ambos conceptos. Ambos conceptos pasarán a tener el mismo rol funcional. Louis sabrá que Clark vuela y que Superman fue criado en Kansas. CLARK estará causalmente relacionado con VUELA y SUPERMAN estará causalmente relacionado con CRIADO EN KANSAS. Lo mismo vale para cualquier otra información asociada a alguno de estos conceptos. De este modo, luego que Louis conoce la identidad, una propuesta de rol funcional pueden explicar por qué los conceptos pasan a ser idénticos. CLARK y SUPERMAN han pasado a ser cognitivamente equivalentes porque tienen el mismo rol funcional. Así, esta propuesta no tiene el problema que tienen las explicaciones a partir de la estructura semántica interna perceptual. Vimos que también podemos evitar de este modo el problema de la estructura semántica no perceptual. Los roles funcionales que determinan el contenido de un concepto no son otra cosa que las relaciones causales sistemáticas que tienen con otras relaciones. Por lo tanto, si dos representaciones difieren respecto de ciertos poderes causales relevantes difieren, *ipso facto*, respecto de su rol funcional constitutivo.

Podemos concluir entonces que podemos ofrecer una caracterización adecuada del contenido conceptual de una estructura *C* en términos de la información en el rol funcional de *C*, *pero que la información que porta C mismo, aunque esta información esté estructurada, no puede identificarse con el contenido conceptual de C*. Repasemos ahora mi propuesta sobre las representaciones perceptuales para determinar, en la sección 4, qué vínculo hay entre el contenido informacional de los conceptos y el contenido informacional de los perceptos.

3. Representaciones perceptuales

En el capítulo pasado argumenté a favor de la caracterización de las representaciones perceptuales a partir de la noción de sistema perceptual. Sin embargo, critiqué varios aspectos de esta propuesta de Prinz y argumenté a favor de otra versión de la misma. En primer lugar rechacé su

caracterización de la noción de sistema perceptual y defendí la principal intuición detrás de las propuestas de la receptividad, la idea de que los sistemas perceptuales son plenamente caracterizados por la función a la que sirven. En segundo lugar, argumenté en contra del vínculo que Prinz establece entre los sistemas y las representaciones perceptuales y propuse un modo alternativo de caracterizar este vínculo. Como vimos, el resultado del enfoque propuesto es que la noción de representación perceptual implica que estas representaciones tienen un contenido informacional. Sin embargo, argumenté también que puede haber estructuras neuronales que porten información pero que no sean representaciones perceptuales. Repasemos brevemente las nociones de sistema perceptual, de representación perceptual y la relación de esta última con la de información.

Si bien Prinz ofrece evidencia en contra de que los sistemas perceptuales sean módulos, considera que hay dos características de los módulos que permiten caracterizar a los sentidos: Sus representaciones y vías neuronales propias. Por otro lado, si bien Prinz rechaza la idea de que los sentidos pueden caracterizarse por cierta pasiva dependencia de inputs externos, la función propia de los sistemas perceptuales, responder a un input específico externo al sistema nervioso, sí permite caracterizar a los sentidos. Argumenté que ni las representaciones ni las vías neuronales sino sólo la función de los sistemas perceptuales son adecuadas para caracterizarlos, porque sólo la caracterización en base a la función específica de los sistemas perceptuales es conceptualmente compatible con el empirismo. Argumenté que, cualquiera sea la posición que uno tenga sobre la tesis empirista, es más conveniente sostener una noción de representación perceptual que sea conceptualmente compatible con ella. Esto es obvio para quienes sostienen, como Prinz que esta tesis es verdadera y para quienes sostienen que es una tesis empírica (que no puede ser falsa sólo por los conceptos que involucra). Pero creo que también es cierto para quienes, como intentaré en este capítulo, argumentan en contra de esta tesis. Pues creo que está en una situación dialécticamente ventajosa quien puede argumentar en contra del empirismo sin emplear una noción de representación perceptual incompatible con esta tesis (es decir, una noción que el empirista también puede aceptar).

De este modo, caractericé a los sistemas perceptuales exclusivamente por medio de su función peculiar y empleé luego esta noción para caracterizar, a su vez, a las representaciones perceptuales. Las representaciones perceptuales no son otra cosa, entonces, que las representaciones que pertenecen un sistema perceptual del modo en que lo he caracterizado. Sin embargo, como señalé en el capítulo pasado, hace falta determinar qué quiere decir “pertenencia”. Como vimos, propuse elucidar sentido adecuado de “pertenencia a un sentido” empleando la nota que propuse como característica de los

sistemas perceptuales, su función específica. Podemos decir que las representaciones perceptuales son las que contribuyen a la función de los sistemas perceptuales. Podemos decir que las representaciones que pertenecen a un sistema perceptual son las representaciones que tienen la función de responder a un input específico externo al sistema nervioso.

Argumenté que este criterio es adecuado pero aún insuficiente. Como vimos, la tesis empirista implica que las representaciones perceptuales sirven tanto a la función de sistemas no perceptuales como a la de los sistemas perceptuales. Esto quiere decir que no podemos llamar a ciertas representaciones “perceptuales” por el hecho de que contribuyen a la función de los sistemas perceptuales y decir que los sistemas no perceptuales emplean representaciones perceptuales. Pues con el mismo criterio podemos decir que las representaciones son “no perceptuales” porque contribuyen a la función de los sistemas no perceptuales y afirmar que los sistemas perceptuales emplean representaciones no perceptuales. Este criterio para caracterizar a las representaciones perceptuales no nos permite distinguir la tesis empirista de que las representaciones perceptuales son empleadas por todos los sistemas cognitivos, de la tesis racionalista de que los sentidos emplean emplean las representaciones no perceptuales de las facultades cognitivas superiores.

Sostuve que, sin embargo, podemos tomar herramientas conceptuales de Prinz mismo para resolver este problema. El punto es simplemente que, si todos los sistemas cognitivos emplean las mismas representaciones, podemos decir que son perceptuales ssi (i) las representaciones en los sistemas no perceptuales son *copias* de representaciones previamente empleadas por los sistemas perceptuales y (ii) el empleo de estas representaciones por parte de los sistemas perceptuales no presupone el proceso de copiado. Las representaciones perceptuales son representaciones que sirven a la función de un sistema perceptual sin la mediación de un proceso de copiado. Por contraste, esto nos da un modo de caracterizar la noción de representación no perceptual. Una representación no perceptual es una representación que es empleada por un sistema no perceptual sin mediación de un proceso de copiado.

Hasta aquí he retomado la caracterización de la noción de representación perceptual en base a la noción de sistema perceptual. Argumenté en el capítulo pasado que esta noción de representación perceptual tiene un vínculo peculiar con la noción de información. Por un lado, ser una representación perceptual implica tener un contenido informacional. Sin embargo, no se da la implicación en la otra dirección: Que una estructura neuronal porte información sobre un hecho externo al sistema nervioso

no implica que sea perceptual.

Hay un vínculo obvio entre la noción de información y la noción de representación perceptual propuesta. Si las representaciones perceptuales tienen la función de correlacionarse con inputs externos al sistema nervioso, portarán entonces información sobre los mismos. Esto es importante, el hecho de que las representaciones perceptuales porten información no es un hecho contingente sino una verdad conceptual. Si la caracterización de la noción de representación perceptual propuesta es adecuada, entonces esta noción implica la de información.

Como afirmé, a primera vista, parece que la implicación también va en la otra dirección. Vimos que la función de los sistemas perceptuales es lo que los distingue de otros sistemas. Pero si sólo las representaciones de los sistemas perceptuales se vinculan con inputs externos entonces sólo las representaciones perceptuales portan información sobre hechos externos al sistema nervioso. Es decir, no sólo ser una representación perceptual implica portar información sino que ser una estructura neuronal que porta información implica ser perceptual. Sin embargo, sostuve que esta última implicación no se da. Una estructura neuronal puede portar información sobre hechos externos al sistema nervioso sin ser una representación perceptual, sin pertenecer a un sistema perceptual.

La razón es que, dada una representación A de un sistema auditivo que responde al patrón de movimiento molecular producido por el ladrido de un perro y la representación V de un sistema visual activada por un patrón de ondas de luz producido por el cuerpo del mismo perro, puede existir también una estructura neuronal N que sea activada tanto por medio de la activación de A como por medio de la activación de V . La estructura N portará información sobre la presencia de un perro, pues sólo será activada en presencia de perros, pero al ser activada por diferentes tipos generales de input proximales no podemos decir que sea perceptual. Pues vimos que las representaciones perceptuales sólo son activadas en presencia de un tipo de input específico. Esto quiere decir que puede haber estructuras neuronales que porten información pero que no sea perceptuales.

Habiendo caracterizado la noción de representación perceptual y determinado su vínculo con el contenido informacional, pasaré a considerar, en la próxima sección, la relación o relaciones entre las representaciones perceptuales y las conceptuales.

4. Lo conceptual y lo perceptual

Sobre la base del enfoque de lo perceptual y lo conceptual que he propuesto en las secciones previas surgen diferentes aspectos de la relación entre estos dos tipos de representaciones. Vayamos considerando cada uno de ellos.

(i) Ser un concepto no implica ser una representación perceptual.

Esto es simplemente un corolario de que la tesis defendida de que la noción de información no implica la de representación perceptual, que puede haber estructuras neuronales que porten información pero no pertenezcan a un sistema perceptual. Veamos este punto con algo de detalle. Vimos que podemos explicar los casos Frege a partir de dos dimensiones informacionales del contenido conceptual. En primer lugar, la información que porta una estructura C constituye el contenido intencional de C . Por otro lado, la información de las estructuras que figuran en el rol funcional de C constituye su contenido cognitivo. Como tanto C como las estructuras en el rol funcional de C pueden portar información sin ser perceptuales, se sigue que lo que hace a C un concepto (su contenido intencional y su contenido cognitivo) no requieren que C sea perceptual.

(ii) Ser una representación perceptual no implica ser un concepto

Vimos que las representaciones perceptuales portan necesariamente un contenido informacional. Ya sabemos que el contenido informacional no es necesariamente perceptual. Pero podemos preguntarnos si es necesariamente conceptual. Si este fuese el caso, entonces ser una representación perceptual implicaría ser un concepto. Creo que este no es el caso. Vimos que el contenido cognitivo de los conceptos requiere que estos estén relacionados causalmente con otras representaciones que porten información. En la sección 2 retome varios argumentos en contra de que la estructura interna de un concepto, sea semántica o sintáctica, pueda dar cuenta de los casos Frege. Mostré que el contenido cognitivo de un concepto dado C debe ser determinado por la información que portan las estructuras con las que esta causalmente relacionada. Si bien la noción de representación perceptual implica que estas mismas representaciones portan información, no implica que deban estar relacionadas con otras estructuras que porten información. La noción de representación perceptual exige que sean empleadas por sistemas perceptuales originalmente, sin la mediación de un proceso de

copia. Pero ser empleado por un sistema perceptual implica sólo correlaciones con inputs externos al sistema nervioso, no implica *necesariamente* relaciones con otras representaciones.

Más aún, supongamos que aceptamos la tesis (en contra de la cual argumenté en la sección 2) de que el contenido cognitivo puede ser explicado por la estructura semántica interna de las representaciones perceptuales, como supone Prinz (2002). Aún si este es el caso, se sigue que ser perceptual no implica ser conceptual. Pues la noción de representación perceptual requiere que estas representaciones porten información sobre inputs externos, pero no implica nada sobre su carácter estructurado o no estructurado. Si, como sostiene Prinz, las representaciones perceptuales tienen una estructura interna, este es un hecho contingente de las mismas, algo que no se sigue de la noción misma de representación perceptual.

Las tesis (i) y (ii) muestran que la tesis empirista no es verdadera en virtud de los conceptos que involucra. No es una verdad conceptual ni que las representaciones perceptuales son conceptos ni que los conceptos son representaciones perceptuales. Pasemos ahora a considerar una tercera tesis:

(iii) No perceptual no implica conceptual

En la sección pasada vimos que una estructura puede portar información sin ser perceptual. Ahora bien, podemos preguntarnos si una estructura que porte información y sea efectivamente no perceptual es necesariamente conceptual. Es claro que dada la noción de contenido cognitivo propuesta, este no es el caso. Una estructura no perceptual puede portar información sin estar relacionada causalmente con otras estructuras que porten información y por lo tanto puede no tener un contenido cognitivo en el sentido propuesto. Así, ser no perceptual no implica ser conceptual. Este tesis nos deja ver un punto importante sobre las representaciones mentales: queda abierta la posibilidad lógica de un tipo de representación informacionalmente caracterizado que no sea ni perceptual ni conceptual. Ahora podemos pasar a la última y más controvertida tesis que quiero sostener en esta sección:

(iv) Conceptual implica no perceptual

Como vimos en (i), ser un concepto no implica ser una representación perceptual. Ahora bien, a primera vista pareciera que ser un concepto tampoco implicaría ser una representación no perceptual.

Pareciera no haber nada en la noción de representación perceptual que impida que posea un rol funcional que involucre otras estructuras que porten información. Sin bien no es necesario que suceda, parece que una representación perceptual *puede* estar relacionada causalmente con otras estructuras que porten información y por lo tanto puede tener un contenido determinado por la información que figura en su rol funcional. Si esto fuese así, entonces pareciera que la tesis empirista, la tesis de que los conceptos empleados por los sistemas no perceptuales son representaciones perceptuales, sería simplemente una tesis empírica, no una verdad conceptual.

Sin embargo, considero que a pesar de esta apariencia inicial, la tesis empirista *no* puede ser verdadera, es conceptualmente falsa (es falsa en virtud de la nociones mismas que involucra). En particular, creo que cuando una representación perceptual adquiere un contenido cognitivo, *necesariamente deja de ser perceptual, pasa a ser una representación no perceptual*. De este modo, si bien es un hecho contingente el que una representación perceptual adquiera un contenido conceptual, esa adquisición la transforma inmediatamente en una representación no perceptual. Por lo tanto, una representación no puede ser al mismo tiempo perceptual y conceptual. Veamos por qué.

Como vimos, la formación del rol funcional de un concepto puede ser caracterizada como la formación de enlaces de memoria entre representaciones. Pero el encargado de la formación de enlaces de memoria no puede ser un sistema perceptual. Como vimos, cada sistema perceptual tiene o manipula representaciones que responden a un tipo de input proximal específico y *no a otro*. Pero los enlaces pueden vincular representaciones de diferentes sistemas perceptuales, representaciones que responden a diferentes tipos de input proximales. Esto es lo que hace, en particular, el tipo de enlace que Prinz (2002) llama “enlace de cohesión” (o “*binding link*”, p.145). Por lo tanto, el sistema que genera nuevas relaciones causales entre representaciones no es un sistema perceptual o parte de un sistema perceptual. Esto quiere decir que la adquisición de un rol funcional y, por lo tanto, la formación de un concepto, depende de un sistema no perceptual. Los conceptos se originan a través del funcionamiento de un sistema no perceptual.

Este hecho nos dice varias cosas importantes. En primer lugar, nos muestra algo general sobre los sistemas cognitivos. Los sistemas cognitivos no sólo pueden llevar a cabo diferentes operaciones con representaciones, manipular representaciones, sino que también pueden *producir* representaciones. Esta tesis es más bien obvia si uno adopta, como lo hice, una propuesta que involucre roles funcionales. Los roles funcionales son conjuntos de relaciones inferenciales o causales entre representaciones que,

al menos en muchos casos, son claramente adquiridos. *Aprendemos*, por ejemplo, que los perros muerden, adquirimos la relación inferencial entre la representación de algo peludo con cuatro patas y la representación del acto de morder. Ahora bien, una vez que notamos que los sistemas cognitivos pueden llevar a cabo estas dos tareas generales podemos descubrir, en segundo lugar, que hay una insuficiencia en la elucidación que ofrecí en el capítulo pasado de la noción de pertenencia a un sistema cognitivo.

Afirmé que una representación pertenece a un sistema cognitivo dado si y sólo si es empleada o manipulada por él sin la mediación de un proceso de copiado. Sin embargo, es claro que ser *producido* por un sistema dado es un criterio de pertenencia a ese sistema al menos igual de fuerte que el criterio de ser empleado o manipulado originalmente por un sistema. Por lo tanto, podemos decir que una representación pertenece a un sistema si y sólo si es o bien producida por ese sistema o bien manipulada por el mismo sin proceso de copiado previo.

Estas consideraciones, junto con el hecho de que los conceptos son producidos por sistemas no perceptuales, implican la tesis de que los conceptos son necesariamente representaciones no perceptuales. Así, si bien las representaciones perceptuales pueden adquirir un contenido conceptual, pueden devenir conceptos, no pueden hacerlo sin dejar de ser representaciones perceptuales. Esto constituye un argumento peculiar en contra del empirismo.

5. Conclusiones

Sobre la base de la elucidación de las nociones de representación perceptual y representación conceptual he ofrecido en esta última sección un conjunto de tesis sobre la relación entre ambas nociones. Las primeras dos tesis establecen que el empirismo, de ser verdadero, es una verdad empírica, no conceptual. He defendido que la noción de concepto no implica la de representación perceptual ni viceversa. En segundo lugar, he mostrado que estas nociones dejan abierta la posibilidad lógica de una representación informacionalmente caracterizada que no sea ni perceptual ni conceptual. Por último defendí la tesis de que no es sólo el caso de que el empirismo no es una verdad analítica, sino que es una falsedad analítica. Ser conceptual implica ser no perceptual. He argumentado que incluso si se da el caso de que las representaciones perceptuales adquieren un contenido conceptual,

esto mismo haría que dejen de ser representaciones perceptuales. Esto quiere decir que hay una distinción excluyente entre representaciones perceptuales y representaciones conceptuales.

En el próximo capítulo tomaré estas nociones no ya para relacionarlas entre sí sino para seguir extendiendo el rango de aplicación del enfoque informacional. Me ocuparé de la noción de conocimiento en general y de la noción de conocimiento perceptual en particular. Argumentaré, en el marco de discusiones recientes en torno a la noción de conocimiento, que este puede ser caracterizado en términos puramente informacionales. En segundo lugar, sostendré que esta caracterización implica que los procesos perceptuales son epistémicamente irrelevantes y que por lo tanto no podemos decir de ningún caso de conocimiento que es un caso de conocimiento perceptual. Es decir, argumentaré en contra de una forma de anti-empirismo epistémico.

Capítulo 6

Conocimiento como información doxástica y un anti-empirismo del conocimiento

La idea de que la noción de información puede servirnos no sólo para caracterizar a la creencia y a la representación perceptual, sino también al conocimiento, es al menos tan vieja como el enfoque informacional mismo. Dretske (1981), desde el título mismo de su obra más influyente, anuncia esta conexión conceptual. En este capítulo presentaré una crítica la concepción informacional del conocimiento de Dretske y voy a ofrecer una concepción informacional alternativa.

Para llegar a ello, sin embargo, voy a empezar con otra discusión en teoría del conocimiento que es aparentemente independiente del enfoque informacional. Se ha desarrollado recientemente una discusión interesante en teoría del conocimiento entre el enfoque de la seguridad y la epistemología de la virtud. La epistemología de la virtud es una familia de teorías cuyo denominador común es la tesis de que para que haya conocimiento es necesario y suficiente que la creencia sea adquirida por medio del ejercicio de una virtud epistémica. Una virtud epistémica es, esquemáticamente, una facultad cognitiva que produce creencias adecuadamente. Algunos ejemplares de epistemología de la virtud son teorías de tipo confiabilista (por ejemplo Axtell (1997), Greco (2002), Baehr (2004), Greco & Turri (2009), Kvanvig (2010), y Turri & Sosa (2010)). En una epistemología de la virtud de tipo confiabilista, “adecuadamente” puede ser entendido, entre otras cosas, como “confiablemente”. Una facultad constituye una virtud epistémica si es confiable, es decir, si produce mayormente outputs verdaderos.

Duncan Pritchard ha defendido y desarrollado una crítica a la epistemología de la virtud por lo menos desde Pritchard (2003) hasta Pritchard (2013). Pritchard ha argumentado repetidamente que la epistemología de la virtud es vulnerable a la suerte epistémica que está en juego en los casos de tipo Gettier, casos en los que el sujeto tiene creencia verdadera justificada pero no conocimiento, dado que

la justificación no basta para que la creencia no sea verdadera sólo por suerte. Esto muestra, según Pritchard, que el conocimiento no puede ser identificado con la creencia verdadera formada por el ejercicio de nuestras facultades epistémicas, que los defensores de la epistemología de la virtud deben incorporar un elemento ajeno a su propuesta para evitar los casos de tipo Gettier. En particular, deben incorporar una condición como la de seguridad, la condición de que la creencia no sea falsa en los mundos cercanos en los que es adquirida. Versiones de la condición de seguridad para el conocimiento fueron ofrecidas por una variedad de autores, incluyendo a Luper (1984; cf. Luper 2006), Sainsbury (1997), Sosa (1999), Williamson (2000), y Pritchard (2002; 2005; 2007a).

En este capítulo argumentaré a favor del enfoque de la seguridad desarrollando una vía argumentativa alternativa a la de Pritchard. En primer lugar, criticaré y rechazaré los argumentos de Pritchard en contra de la epistemología de la virtud y a favor de la condición de seguridad y ofreceré un par de argumentos alternativos. Considero que la defensa de Pritchard del enfoque de la seguridad no es adecuado pero que sin embargo pueden construirse argumentos alternativos que muestran que la condición de seguridad es necesaria y suficiente para el conocimiento.

En segundo lugar, mostraré que el enfoque informacional de Dretske es vulnerable a la misma crítica que padece la epistemología de la virtud. La propuesta de Dretske es vulnerable a la suerte epistémica y por lo tanto requiere ser suplantada por la noción de seguridad. Sin embargo, considero que es posible proponer una noción informacional equivalente a la noción de seguridad. Esto constituye una ventaja ya que hace que mi propuesta cuente con un enfoque que unifica conceptualmente a las nociones de contenido conceptual y justificación doxástica. Ambas pueden ser caracterizadas en términos de propiedades informacionales.

Por último, sostendré que, como esta concepción del conocimiento es incompatible con pensar que los procesos cognitivos son epistémicamente relevantes, tenemos que negar que haya propiamente *tipos* de conocimiento. En particular, debemos negar que haya conocimiento perceptual. Esto constituye una forma radical de anti-empirismo epistémico en la medida en que no se niega meramente que todo conocimiento sea perceptual, sino que algún caso de conocimiento lo sea.

Antes de comenzar con la discusión propiamente epistemológica, hace falta aclarar ciertos aspectos de la noción de contenido informacional y su relación con la noción informacional de justificación.

1. Dos dimensiones informacionales de la creencia: Contenido y justificación

Estuve retrasando varios capítulos una cuestión importante dentro del marco de una teoría informacional de la mente. Vimos que decir que una estructura porta información sobre un hecho quiere decir que dicha estructura se encuentra en una relación nomológica o de dependencia contrafáctica con dicho hecho. Una estructura C porta información sobre la presencia de una propiedad P sólo si la instanciación de C no se hubiese dado si P no se hubiese instanciado, es decir, si C no se instancia en un mundo cercano en el que P no lo haga. Así, la noción de portar información se refiere a una instancia particular de C que se activa en la presencia de P y que no se hubiese activado en su ausencia. Ahora bien, identificar al contenido de C con la información que porta es incompatible con el hecho de que C puede ser aplicado erróneamente, esto es, con el hecho de que pueda instanciarse en ausencia del objeto o propiedad al que refiere. De este modo, no podemos identificar al contenido de C con la información que porta.

Este constituye uno de diversos problemas que se le han presentado a la propuesta informacional y a las teorías causales del contenido mental en general. Como dije en el capítulo 2, dado que hay una extensísima literatura en torno a estos problemas y dado que coincido con Fodor (2008) en que estos problemas no son insolubles (aunque tal vez no se haya llegado a una solución definitiva aún), no tomaré partido sobre una solución particular a todos estos problemas. Sin embargo, ofrecer una solución al problema recién mencionado, el problema de explicar informacionalmente la aplicación errónea de un concepto, es crucial para dar una teoría informacional del conocimiento. Después de todo, si voy a ofrecer una teoría informacional de la condición de justificación para el conocimiento, debo poder distinguirla de la teoría informacional de la condición de verdad. Si bien se ha discutido la independencia de estas dos condiciones (por ejemplo, desde una perspectiva coherentista de la verdad), la negación de esa independencia es al menos controvertida y por lo tanto considero preferible una propuesta que no la presuponga.

Para llevar a cabo esta distinción me basaré en la propuesta clásica de Dretske (1981). Dretske considera que el contenido de una instancia c del concepto C no está determinado por la información que porta c sino por la información que las instancias de C tienen la función de portar (p. 193). Dretske propone una analogía entre los conceptos y los diferentes rasgos de un mapa. Una línea azul en un mapa tiene la función de indicar la presencia de un río, de portar esa información. Esta es una función

asignada por una convención. Esta función es lo que hace posible que una línea represente erróneamente un río cuando, al no haber un río presente en la realidad, la línea no porta efectivamente información sobre su presencia. Dretske propone que la propiedad sobre la que una estructura neuronal tiene la función de portar información es la propiedad a la que empezó a responder en el momento de aprendizaje, en el momento de adquisición del concepto constituido por esa estructura. Es decir, un concepto *C* tiene la función de portar información sobre una propiedad *P* ssi las primeras instancias de *C* portaban efectivamente información sobre *P*. De este modo, aunque instancias posteriores de *C* no porten información sobre *P* aún representarán a *P*. En estos casos las instancias de *C* representarán erróneamente a *P* (p. 193).

Esto quiere decir que podemos interpretar la idea de contenido informacional que he empleado para desarrollar mi propuesta sobre conceptos según estas líneas. Cuando digo que el contenido de un concepto *C* está determinado por su propio contenido informacional y por el contenido informacional de las estructuras en su rol funcional quiero decir que el contenido de *C* está determinado por la información que todas estas estructuras tienen la función de portar, la información portada por las primeras instancias de todas esas estructuras (o por la información que portaban esas estructuras las primeras veces que se activaron).

Hay respuestas alternativas al problema del error para propuestas informacionales o causales y han sido ofrecidas diferentes versiones de la propuesta de Dretske. Una de las propuestas alternativas más conocidas es la de Fodor (por ejemplo, Fodor 2008, pp. 204-205). Sin embargo, dado que es un tema muy extenso para ser tratado exhaustivamente aquí y que sólo necesito esta noción para trazar la distinción entre contenido informacional e información de una estructura, remito a la argumentación de Prinz a favor de la propuesta recién expuesta (2002, pp. 241-261). El enfoque de Prinz es ligeramente diferente en aspectos que son relevantes para la caracterización informacional del contenido pero no ahondaré en ellos aquí dado que no son relevantes para la distinción que me interesa.

En este capítulo defenderé la tesis de que el conocimiento es la creencia que porta información sobre el hecho conocido. La creencia *C* de un sujeto *S* de que hay un perro presente constituye conocimiento sii, hay un perro presente y *C* porta la información de que hay un perro presente. Así *c*, una instancia de *C*, es una creencia caso de que hay un perro presente porque las primeras instancias de *C* portaban esta información, pero constituye conocimiento porque *c* mismo porta esta información.

En lo que sigue ofreceré una extensa defensa de esta caracterización informacional del

conocimiento. En la próxima sección criticaré los argumentos de Pritchard en contra de la epistemología confiabilista de la virtud y a favor de una propuesta de la seguridad y ofreceré argumentos alternativos. En la sección 3 argumentaré que la propuesta informacional de Dretske adolece el mismo problema que las epistemologías de la virtud y que la propuesta informacional que propongo allí es equivalente al enfoque de la seguridad y por lo tanto, los argumentos a favor de este último valen a favor de aquella. Por último, argumento que esta noción de conocimiento, unida a la caracterización ofrecida en el capítulo 4 de los sistemas perceptuales implica que ningún caso de conocimiento es un caso de conocimiento perceptual. Argumento a favor de una forma de anti-empirismo respecto del conocimiento.

2. El enfoque de la seguridad

Recientemente diferentes autores han defendido un enfoque epistemológico llamado “epistemología de la virtud”. El denominador común de estas propuestas es que el factor constitutivo de la justificación doxástica y el conocimiento son las virtudes epistémicas, entendidas, de modo muy general, como las facultades cognitivas que producen adecuadamente outputs verdaderos. Algunos ejemplares de este tipo de propuesta (por ejemplo, Axtell (1997), Greco (2002), Baehr (2004), Greco & Turri (2009), Kvanvig (2010), y Turri & Sosa (2010)) constituyen una variedad de confiabilismo. En una epistemología de la virtud de tipo confiabilista, “adecuadamente” puede ser entendido, entre otras cosas, como “confiablemente”. Me interesa en esta sección examinar una línea de crítica que se le ha opuesto a este tipo de confiabilismo y que se ha sostenido hasta la actualidad, a mi parecer, sin respuesta adecuada. El resultado de esta crítica es lo que podemos llamar el “enfoque de la seguridad” y que, como argumentaré en la sección III, es equivalente a la propuesta informacional que quiero defender.

Duncan Pritchard ha venido defendiendo y desarrollando una crítica a la epistemología de la virtud por lo menos desde Pritchard (2003) hasta Pritchard (2013). Pritchard ha argumentado repetidamente que la epistemología de la virtud es vulnerable a la suerte epistémica que está en juego en los casos de tipo Gettier, casos en los que el sujeto tiene creencia verdadera justificada pero no conocimiento, dado que la justificación no basta para que la creencia sea verdadera sólo por suerte. El conocido caso de Chisholm (1977, p. 105) de las ovejas funciona en este sentido. En este caso el sujeto

adquiere la creencia de que hay una oveja en el campo empleando sus capacidades perceptuales confiables y por lo tanto, sus capacidades epistémicas. Más aún, su creencia es verdadera. Sin embargo, la creencia no constituye conocimiento ya que lo que está viendo el sujeto no es una oveja sino un objeto con forma de oveja que está tapando una oveja detrás de él. La razón por la que el sujeto carece de conocimiento es porque la creencia es sólo accidentalmente verdadera. La creencia podría fácilmente haber sido falsa si la oveja verdadero hubiese estado pastando en otro campo.

Como el caso muestra, según Pritchard, que el conocimiento no puede ser identificado con la creencia verdadera formada por el ejercicio de nuestras facultades epistémicas, los defensores de la epistemología de la virtud deben incorporar un elemento ajeno a su propuesta para evitar los casos de tipo Gettier. En particular, deben incorporar una condición como la de seguridad, la condición de que la creencia no sea falsa en los mundos cercanos en los que es adquirida (Luper (1984) y (2006), Sainsbury (1997), Sosa (1999), Williamson (2000), y Pritchard (2002; 2005; 2007a)). Lo que según Pritchard muestran casos como este es que, en contra de la epistemología de la virtud, la posesión o falta de conocimiento depende frecuentemente de variables ajenas al funcionamiento adecuado de las facultades cognitivas del sujeto cognoscente. Pritchard llama a este fenómeno “dependencia epistémica”.

Muchos epistemólogos de la virtud han defendido la idea de que este enfoque es perfectamente capaz de lidiar con el problema de la dependencia epistémica sin adoptar la condición de seguridad o alguna otra condición semejante, ajena al enfoque. Tal abordaje es encontrado en autores como Ernest Sosa (1988; 1991; 2007; 2009), Linda Zagzebski (1996; 1999) y John Greco (2003; 2007; 2008; 2009a; 2009b; 2009c). Si bien las estrategias de estos autores para lidiar con los casos Gettier difieren en aspectos relevantes, comparten la idea de que el conocimiento no es meramente el éxito cognitivo (la verdad de la creencia) más el ejercicio de la capacidad epistémica relevante, sino el éxito epistémico *a causa* del ejercicio de la facultad epistémica relevante. Por ejemplo, en el caso de la oveja, el hecho de que la creencia del sujeto sea verdadera no se debe al ejercicio de sus facultades epistémicas sino al hecho accidental de que la oveja está detrás de un objeto con forma de oveja percibido por medio dichas facultades. Esta propuesta mostraría que en el caso de la oveja no se da el fenómeno de dependencia epistémica, como piensa Pritchard, porque la falta de conocimiento del sujeto se debe a que sus facultades cognitivas no operan adecuadamente.

Ahora bien, Pritchard considera que hay casos que pueden volver a generar el mismo problema

a esta propuesta modificada de la epistemología de la virtud. Cree que la dependencia epistémica puede darse para casos en los que se cumple la condición propuesta en el párrafo anterior. Pritchard propone un caso análogo al de las tierras gemelas de Putnam. Se trata de el caso de las tierras epistémicamente gemelas. Dicho brevemente, la tierra epistémicamente gemela es una tierra exactamente igual a la nuestra pero en la que tenemos tanto muestras de H_2O como muestras de xyz (un compuesto químico cuya apariencia y propiedades causales relevantes para nuestra interacción cotidiana son idénticas a las del H_2O). Supongamos que un sujeto *S* en la tierra gemela epistémica genera la creencia verdadera de que está viendo agua ejerciendo su percepción visual en condiciones normales de iluminación, distancia, etc. En este caso, a diferencia del caso de la oveja, como la creencia de *S* es el resultado de una interacción causal entre una muestra de H_2O y las facultades perceptivas de *S*, la creencia es verdadera *a causa* del ejercicio de sus facultades. Parece que, sin embargo, *S* no tiene conocimiento ya que su creencia podría haber sido fácilmente falsa. Hay situaciones contrafácticas cercanas donde hubiese formado la misma creencia pero a partir de percibir xyz (Pritchard 2013, p. 10).

Considero que este caso es exactamente igual, en los aspectos relevantes para lo que Pritchard quiere mostrar, al caso de Goldman (1976) de las fachadas de granero. En este caso hay también un sujeto percibiendo un granero real en condiciones perceptuales normales y formando la creencia verdadera de que hay un granero presente. La creencia es verdadera a causa del ejercicio de sus facultades perceptuales. Pero el sujeto está ubicado en un entorno en el que hay muchas meras fachadas de granero, objetos que no son graneros pero que son perceptualmente indiscernibles por el sujeto de graneros reales. Esto hace que, aunque el sujeto adquiere la creencia verdadera a causa de sus facultades epistémicas en condiciones favorables, haya situaciones contrafácticas cercanas en las que el sujeto adquiere la misma creencia pero esta resulta falsa y por lo tanto, parece que no tiene conocimiento. Esto mostraría, según Pritchard, que la versión mejorada de la epistemología de la virtud es aún vulnerable al fenómeno de la dependencia epistémica. Aunque se adquiriera una creencia verdadera *por medio* del funcionamiento adecuado de una facultad cognitiva, puede haber factores ajenos a esas facultades que impidan que el sujeto tenga conocimiento.

Ahora bien, John Greco, uno de los defensores de la versión revisada de la epistemología de la virtud, ha respondido al argumento de Pritchard afirmando que el sujeto de la tierra gemela epistémica no tiene la habilidad epistémica en la situación mencionada ya que esta requiere de la confiabilidad de la facultad ejercida, y esta confiabilidad es aquello de lo que carece la facultad en el entorno en el que se encuentra (2009c, 21-22). Lo mismo vale para el caso de Goldman. A esto Pritchard responde

diciendo que el sujeto no carece de la habilidad si consideramos que lo que hace falta es sólo que la facultad que ejerce sea confiable en un entorno más global. Supongamos que las muestras de xyz se encuentran *sólo* en el entorno inmediato del sujeto, que no hay muestras de esta sustancia en otros entornos de la tierra gemela epistémica. Si esto es así, la habilidad del sujeto de identificar perceptualmente muestras de agua es confiable en el entorno más global que es la tierra gemela epistémica. Lo mismo podemos decir del caso de Goldman, los pueblos minados con fachadas de granero no son algo común. En general identificar graneros a partir de la percepción de su fachada es un procedimiento confiable. Sólo en el entorno local carece de confiabilidad. De este modo, Pritchard diría que el sujeto del caso de Goldman, como el de su propio caso, no deja de tener la habilidad epistémica relevante en el entorno descrito (Pritchard 2013, p. 14) .

Considero que la respuesta de Pritchard a Greco tiene un problema. No podemos decir que las facultades cognitivas en los ejemplos de Pritchard y Goldman son confiables porque en un contexto más global no se dan los factores de la situación actual y por lo tanto producen mayormente creencias verdaderas. Pues si decimos esto, tendremos que decir que las facultades cognitivas son confiables en contextos locales en los que claramente no lo son. Los casos en los que típicamente diríamos que la adquisición perceptual de creencias no es confiable son los casos en los que no se dan ciertas condiciones normales del entorno que requiere la formación confiable de experiencias perceptuales. Por ejemplo, en el caso de la percepción visual, si no hay iluminación o distancia adecuadas respecto del objeto, o si el objeto está ocluido por otro en el campo visual (este es el caso mencionado de la oveja) etc., la facultad visual ya no es confiable, aún cuando la creencia perceptual sea verdadera. Ahora bien, al igual que los factores de los ejemplos que consideramos (fachadas de granero y muestras de xyz), estos factores no impiden que la facultad sea confiable en un contexto más global, ya que estos son factores que no se dan en este contexto más amplio. La percepción es una facultad confiable en una amplia gama de contextos en los que tenemos iluminación adecuada, distancia y orientación adecuadas, etc. Pero esa mera confiabilidad global no hace que la facultad sea confiable en el contexto local en el que sí se dan estos factores. Si no suficiente luz, la percepción visual simplemente no es confiable. Por lo tanto no es la mera confiabilidad global lo que hace que la percepción sea confiable, en los ejemplos de Pritchard y Goldman, porque esto significaría que la percepción sería confiable en contextos en los que claramente no lo es.

Creo que la respuesta de Pritchard a Greco sería más plausible si se concentrase no en la confiabilidad global, sino en la relación de los factores de su ejemplo con el funcionamiento de las

facultades cognitivas. A primera vista, los factores de los ejemplos de Pritchard y Goldman, a diferencia de factores como la iluminación y distancia respecto del objeto percibido (los factores para el funcionamiento normal de la percepción visual) *son ajenos a las capacidades cognitivas del sujeto, no influyen específicamente sobre su confiabilidad*. La diferencia entre estos dos tipos de factores es que las condiciones para la formación de experiencias visuales son condiciones normales del funcionamiento confiable de cierta facultad cognitiva. Por el contrario, los factores del entorno que describen los casos de Pritchard y Goldman (muestras de xyz y fachadas de granero) parecen no ser factores de este tipo, no parecen constituir condiciones normales del funcionamiento de una facultad. La presencia fachadas de granero o muestras xyz en la situación actual no parecen afectar la percepción visual o ninguna otra facultad relevante. De este modo, considero que es esta característica la que muestra que el caso de Pritchard es un caso genuino de dependencia epistémica. Los factores que impiden que el sujeto tenga conocimiento no constituyen condiciones normales o anormales para el funcionamiento adecuado de la facultad cognitiva que el sujeto emplea.

Ahora bien, creo que esta consideración sobre los casos de Pritchard y Goldman, si bien más adecuada que la noción de confiabilidad global, es aún errada. Considero que los factores que en estos ejemplos impiden que los sujetos tengan conocimiento pueden caracterizarse como factores cognitivos, factores que contribuyen a la confiabilidad de una facultad cognitiva. En particular, creo que esto se sigue si consideramos que la confiabilidad del proceso de formación de una creencia perceptual no depende sólo de los factores relevantes para la formación confiable de la experiencia (como iluminación, distancia, etc.), sino también los factores relevantes para *la formación de una creencia perceptual a partir de la experiencia*.

Por ejemplo, el sujeto de Goldman basa su creencia en la presencia de un granero sólo en la percepción de la fachada del mismo. *Esta parte de su proceso perceptual* (que determinadas notas de la experiencia sean suficientes para disparar la creencia) deja de ser confiable en el pueblo de las fachadas. Si estuviese en un entorno sin meras fachadas esta parte de su proceso perceptual de formación de creencias sería confiable. El proceso perceptual de formación de creencia no es confiable en el pueblo de las fachadas. Por el contrario, si el sujeto de Goldman formase su creencia en la presencia de un granero a partir de una percepción más minuciosa del mismo, entonces el proceso sería confiable aún en el pueblo de los graneros. Lo mismo podemos decir del caso de Pritchard. La presencia de muestras de xyz en el entorno inmediato del sujeto de la situación de Pritchard afecta la confiabilidad de la adquisición de creencias en la presencia de agua *a partir de la experiencia de las*

propiedades superficiales del agua. Aquí también las propiedades del entorno son relevantes para la confiabilidad de una parte del proceso cognitivo del sujeto. La diferencia de este caso con el de Goldman es que aquí parece que no disponemos de un modo de mejorar nuestra experiencia visual de la muestra percibida de modo tal que esta se vuelva confiable en el contexto local. Una inspección más minuciosa requeriría recurrir a instrumentos para determinar la estructura atómica de la sustancia percibida. Pero aún es el caso de que *esta falta de confiabilidad puede describirse en relación con las características del proceso cognitivo que forma la creencia.* La creencia no es confiable porque *la experiencia no es suficientemente discriminatoria, en relación con la creencia que se quiere formar.* La apariencia superficial del agua no me permite discriminar agua de xyz. Esto quiere decir que los factores del entorno que en los ejemplos de Pritchard y Goldman impiden el conocimiento pueden ser caracterizados como factores que no afectan meramente la confiabilidad de la creencia sino también de las facultades cognitivas que la producen. Considero que esta respuesta a la crítica de Pritchard es más adecuada que la de Greco (quien afirma meramente que nuestras facultades no son confiables en el ejemplo) porque explica *exactamente por qué* podemos afirmar que nuestras facultades cognitivas no son confiables en la situación descripta.

Por este motivo, considero que el ni caso de Pritchard ni el clásico de Goldman funcionan para minar una propuesta de la epistemología de la virtud. Casos como los de Pritchard y Goldman no funcionan como contraejemplos a la suficiencia de la confiabilidad de las facultades cognitivas para el conocimiento. Sin embargo, considero que hay otro tipo de casos que sí pueden servir para minar la suficiencia. Curiosamente, Goldberg (2010) ofrece un caso de este tipo pensando que sostiene justamente una intuición contraria. Sammy es un fanático de los Yankees que sigue obsesivamente los resultados de los juegos en la sección deportiva del diario. Sammy odia cuando los Yankees pierden. Cada vez que eso sucede, forma una creencia sobre el resultado del juego basado en sus deseos. En estos casos cree que los Yankees ganaron, cualquiera sea la evidencia que tenga a favor de que perdieron. Sin embargo, cuando los Yankees ganan, forma la creencia de que lo hicieron basado en el reporte confiable del diario. Una mañana un extraño que no sabe nada sobre Sammy le pregunta si los Yankees ganaron y Sammy responde que sí. Como el foco de Goldberg (2010) es el conocimiento testimonial, no se pregunta directamente si la creencia de Sammy es confiable sino si su testimonio lo es. Goldberg responde que esto dependerá del proceso que haya producido el testimonio. Considera que diríamos que el testimonio es confiable si es producido por la aceptación del reporte del diario y no confiable si se basa en sus deseos de que los Yankees ganen. Ahora bien, Goldberg argumenta que

cuando diríamos que Sammy es un testigo confiable esto es porque también diríamos que es un creyente confiable y no hay diferencia notable entre los casos de confiabilidad.

Por el contrario, argumentaré que como intuitivamente diríamos que no es un testigo confiable y no hay diferencia entre los casos de confiabilidad, tampoco podemos decir que sea un creyente confiable. Claramente, nadie que sepa la condición de Sammy estaría dispuesto a pedirle testimonio o a atribuirle conocimiento testimonial a alguien que haya adquirido la creencia sobre el resultado del partido a partir de su testimonio. *Pase lo que pase, Sammy nos dirá que los yankees han ganado.* Pero la condición de Sammy no involucra nada que tenga que ver con su condición específicamente como testimoniante. El es sincero, expresa confiablemente sus creencias. Su condición afecta específicamente su formación de creencias y, por extensión, afecta su confiabilidad testimonial. Por lo tanto, si no lo tomaríamos por un testigo confiable, como creo que es claro, tampoco podemos considerarlo un creyente confiable. Ahora bien, si Sammy no es un creyente confiable y por lo tanto no tiene conocimiento, según los estándares confiabilistas, entonces la confiabilidad de la facultad cognitiva de formación de creencias no basta para tener conocimiento. De modo general, todo sujeto que emplee una facultad cognitiva confiable para formar una creencia pero que emplee una facultad alternativa no confiable en una situación contrafáctica cercana, constituirá un caso en contra de la suficiencia de la confiabilidad de las facultades cognitivas para el conocimiento. Considero que este tipo de casos funcionan mejor que el de Pritchard en contra de la suficiencia de la condición de la epistemología de la virtud para el conocimiento.

Ahora bien, Pritchard argumenta no sólo contra la suficiencia sino también en contra de la *necesidad* de las habilidades epistémicas para obtener conocimiento. Pritchard toma los casos ordinarios de conocimiento testimonial para mostrar que el ejercicio de las habilidades epistémicas de un sujeto no son necesarias para la adquisición de conocimiento. La idea de Pritchard es que en los casos ordinarios de conocimiento testimonial, si bien hay ciertas capacidades epistémicas por parte del sujeto que adquiere conocimiento (como registrar que no hay evidencia en contra de la confiabilidad del testimoniante e interpretar adecuadamente sus palabras) el grueso de la habilidad epistémica que hace a la confiabilidad de la creencia testimonial está dada por los procesos cognitivos del informante, los procesos cognitivos por los que este puede transmitir confiablemente la información recibida por el sujeto cognoscente.

No veo que este tipo de casos pueda cumplir la función que Pritchard quiere asignarles. En

primer lugar, es claro que si se concede, como Pritchard lo hace, que hay algún tipo habilidad cognitiva en juego en la formación de la creencia testimonial, entonces el caso del conocimiento testimonial no puede mostrar que el ejercicio de habilidades cognitivas no es necesario para este tipo de conocimiento. En segundo lugar, aún si mostrase que hay casos de conocimiento que no requieren habilidades cognitivas por parte del sujeto cognoscente, el epistemólogo de la virtud podría responder que sí son necesarias habilidades cognitivas en la mente del testimoniante y que esto basta para la epistemología de la virtud. La tesis central de Goldberg (2010) es justamente que para muchos casos de conocimiento (todos los casos de conocimiento testimonial) la confiabilidad de las facultades epistémicas en sujetos *diferentes del sujeto cognoscente* son necesarias. Goldberg considera que esta es una forma de confiabilismo anti-individualista. La epistemología de la virtud podría tomar esta perspectiva anti-individualista para los casos de conocimiento testimonial. Si aceptamos esta perspectiva, no necesitamos renunciar a la tesis de que la confiabilidad de las facultades cognitivas es necesaria para el conocimiento.

Sin embargo, al igual que con el argumento anterior, considero que Pritchard ha errado meramente con el caso, no con la tesis que quiere defender. Para defender esta tesis, propondré un híbrido del caso de Bertha (llamado “mentiroso consistente”) ofrecido por Lackey (2008) y del caso de las fachadas de Goldman (1976). Antes de presentar el caso, quisiera hacer algunas consideraciones críticas sobre el caso de Lackey. Lackey (2008) ofrece el caso de Bertha para defender una de sus tesis centrales sobre el conocimiento testimonial: Que la confiabilidad de la creencia no es necesaria para la confiabilidad del testimonio. Veamos de qué se trata. Bertha es una joven que como resultado de una lesión cerebral tiene una disposición a mentir respecto de sus percepciones visuales respecto de animales salvajes. Sus padres deciden llevarla al neurocirujano, quien cree que es necesario operarla pero que se da cuenta al mismo tiempo que la lesión no puede revertirse. Por lo tanto, decide generar una segunda lesión que produce fallas sistemáticas en su adquisición de creencias a partir de experiencias visuales de animales salvajes. Como resultado de esta lesión, Bertha se convierte en una creyente radicalmente *no confiable* pero consistente, respecto de sus experiencias de animales salvajes y en una testimoniante *radicalmente no sincera* pero muy consistente, respecto de esta información. Por ejemplo, cada vez que ve un ciervo cree que vio un caballo y testifica no sinceramente que vio un ciervo. Cuando ve una jirafa cree que ve un elefante y testifica haber visto una jirafa, y así sucesivamente. Por lo tanto, la lesión que produjo el médico, convierte a Bertha en una testigo confiable respecto de su percepción de animales salvajes. (Lackey 2008, pp. 53 y 54).

Lackey considera que este caso muestra que la confiabilidad del proceso que produce la creencia es independiente de la confiabilidad del proceso que produce el testimonio (Lackey 2008, p. 55). Por su parte, Goldberg (2010) concede este punto, pero argumenta que esto no implica que la confiabilidad del testimonio no dependa de la confiabilidad del proceso cognitivo que lo produce. Es decir, según Goldberg, este caso no muestra que el testimonio puede ser confiable a pesar de que parte del proceso que lo produce no lo sea, sino simplemente que el proceso que produce a la creencia y el que produce el testimonio son diferentes. El proceso que produce el testimonio de Bertha es un proceso cognitivo que tiene su testimonio como output y que resulta, luego de la segunda lesión, altamente confiable, aún cuando el proceso de formación de creencias perceptuales que media la producción del testimonio no lo sea (Goldberg 2010, p. 27). Lo que a Goldberg le interesa mostrar entonces, es que el caso de Lackey no muestra que el testimonio pueda ser confiable sin que el proceso cognitivo o facultad cognitiva que lo produce lo sea.

Sin embargo, creo que si reparamos en que el proceso que produce el testimonio de Bertha es un proceso que tiene lo que llamaré “confiabilidad condicional”, seremos llevados a una conclusión diferente. La distinción entre procesos dependientes de creencia y procesos independiente de creencia es una distinción propuesta por Goldman (1976) y pretende usarse para mostrar que la confiabilidad de un proceso no debe evaluarse siempre de la misma manera. Los procesos dependientes de creencia, como las inferencias, son procesos cuyos inputs son creencias. Los procesos independientes de creencia, como la percepción, son procesos cuyos inputs no son creencias. La diferencia entre estos tipos de proceso es que la confiabilidad de un proceso dependiente de creencia no es una función sólo de la confiabilidad del proceso, es decir, de la confiabilidad del proceso dado que su input sea verdadero, sino también *de la confiabilidad del proceso que forma el input*. Esto es claro en el caso de una inferencia. Aún si la inferencia de una creencia a otra es altamente confiable, si la creencia original fue formada de modo no confiable (por ejemplo, sobre la base de mis deseos) entonces el proceso que forma la creencia no será confiable. Es claro que esta noción de dependencia de creencia puede extenderse para ser aplicada a procesos que tengan como inputs no sólo creencias sino estados con contenido de cualquier tipo. Por ejemplo, si consideramos que las experiencias perceptuales no son creencias pero tienen un contenido que puede ser evaluado en virtud de su corrección o incorrección entonces podemos decir que el proceso de formar una creencia a partir de una experiencia es un proceso cuya confiabilidad es, al igual que los procesos inferenciales, dependiente de la confiabilidad del input. Goldberg (2010), argumentando que el proceso de adquisición de conocimiento testimonial

es dependiente en este sentido acuña la expresión “proceso dependiente de cuasi-creencia”. Considero que la expresión “confiabilidad condicional” es preferible ya que no nos compromete con la tesis de que cualquier estado con contenido, que sea pasible de evaluación semántica, sea una “cuasi-creencia” o no nos obliga a ofrecer una elucidación de esta noción (Goldberg mismo no intenta ofrecerla).

Ahora bien, considero que es claro que el proceso cognitivo de Bertha es un proceso que tiene no sólo como output, sino también *como input* un estado que puede ser evaluado en función de su confiabilidad, es decir, es un proceso de confiabilidad condicional. Bertha produce un testimonio a partir de cierta creencia. Más aún, la creencia que funciona de input resulta de un proceso que a su vez tiene un input que puede ser evaluado en función de su confiabilidad: la experiencia perceptual. Así, el testimonio de Bertha depende de tres procesos cognitivos conectados que pueden ser evaluados independientemente respecto de su confiabilidad: el proceso de producción de una experiencia perceptual, el proceso de producción de una creencia a partir de esa experiencia y el proceso de producción de un testimonio a partir de la creencia. El proceso de producción del testimonio de Bertha es un *proceso cognitivo complejo* compuesto de, constituido por, tres procesos diferentes cuya confiabilidad puede ser evaluada independientemente.

Considero que en la situación descrita por Lackey, dos de los tres procesos de Bertha no son confiables. El proceso de producción de experiencias visuales en Bertha es confiable, pero la producción de creencias a partir de esas experiencias y la producción de testimonio a partir de creencias no lo son. Estos últimos dos procesos son procesos de confiabilidad condicional. Son confiables ssi producen creencias mayormente verdaderas cuando sus inputs son verdaderos (como en el caso de las inferencias o la memoria). Pero es claro que los dos procesos de Bertha no son condicionalmente confiables en este sentido. Cuando las experiencias de animales salvajes de Bertha sean verdaderas o adecuadas, sus creencias no lo serán. Cuando las creencias de Bertha sean verdaderas, sus testimonios no lo serán.

Goldberg considera que el caso de Bertha muestra que el proceso cognitivo que produce el testimonio es confiable aunque el que produzca la creencia no lo sea. Lo que pierde de vista este análisis es que el proceso de formación de la creencia es *parte* del proceso de producción del testimonio. El proceso total de producción del testimonio de Bertha es un proceso complejo cuya confiabilidad, de acuerdo con el criterio del confiabilismo de procesos mismo, debe ser evaluada en función de la confiabilidad (condicional o no) de los procesos que lo componen. Lo que creo que sí

muestra es caso de Bertha es que la confiabilidad del testimonio puede darse aún cuando los procesos cognitivos que lo producen no sean confiables. Por lo tanto, en contra de Goldberg, la confiabilidad del testimonio no puede ser entendida en función de la confiabilidad de estos procesos cognitivos que lo producen.

Ahora bien, considero que si construimos un análogo del caso de Bertha para el conocimiento perceptual, un híbrido del caso de Bertha y el de Sam (el sujeto de las fachadas de granero) podemos mostrar, en contra de la epistemología de la virtud, que la confiabilidad de las facultades epistémicas no es necesaria para el conocimiento. Supongamos que un sujeto, Sam, posee una falla sistemática en sus facultades visuales: cada vez que mira un granero, tiene la experiencia de estar frente a un elefante y viceversa. Sam tiene también una falla sistemática en la formación de sus creencias perceptuales: Cuando tiene la experiencia de estar frente a un elefante forma la creencia de que está frente a un granero y viceversa. Así, el proceso de Sam de formación de creencias sobre la presencia de elefantes o graneros está constituido por dos procesos ninguno de los cuales es confiable (al igual que el proceso de formación de testimonios de Bertha). La percepción visual de elefantes y graneros arroja sistemáticamente outputs falsos. La formación de creencias perceptuales sobre la presencia de graneros y elefantes es un proceso de confiabilidad condicional que arroja sistemáticamente outputs falsos, dados inputs verdaderos y confiables. Sin embargo, Sam forma confiablemente creencias verdaderas sobre la presencia de elefantes y graneros. De este modo, el ejercicio de habilidades epistémicas, donde esto involucra, al menos, el ejercicio de facultades o procesos cognitivos confiables, no es necesaria para adquirir conocimiento.

Los argumentos que ofrecí, a partir de los casos de Sammy y de Sam, en contra de la necesidad y de la suficiencia de la confiabilidad de los procesos cognitivos que forman una creencia para que esta constituya conocimiento parecen apuntar a favor de una condición como la de seguridad de Pritchard. Lo que falla en el caso de Sammy, a pesar de que se da la confiabilidad de la facultad cognitiva, y lo que se da en el caso de Sam, a pesar de la falla de la facultad cognitiva, es la seguridad de la creencia. En el primer caso, a pesar de la confiabilidad de la facultad cognitiva, la creencia es falsa en situaciones contrafácticas cercanas y en el segundo caso, a pesar de la no confiabilidad de la facultad cognitiva, la creencia no sería falsa en situaciones contrafácticas cercanas. Parece entonces que lo único necesario y suficiente para el conocimiento parece ser el valor de verdad en situaciones contrafácticas cercanas de la creencia adquirida en la situación actual.

3. La teoría informacional del conocimiento

Como dije al comienzo del capítulo, la idea de proponer una teoría informacional del conocimiento es parte del proyecto informacional de Dretske (1981). La tesis de Dretske es que un sujeto S sabe que p si la creencia de S de que p es causada (o causalmente sostenida) por la información de que p (p. 86). Dretske desarrolla su propuesta del siguiente modo. Supongamos que una señal s porta la información de que r es F y la porta virtud de tener cierta propiedad G . No cualquier golpe en la puerta le dice al espía que el mensajero ha llegado. La señal son tres golpes rápidos seguidos de una pausa y luego tres golpes más. Es esta secuencia específica de golpes la que porta la información de que el mensajero ha llegado. No es, por ejemplo, el tono de los golpes o el momento en el que ocurren lo que es significativo. Es el patrón temporal de golpes lo que constituye la propiedad G que porta la información. Cuando una propiedad como G porta la información de que r es F entonces Dretske dice que la información de que r es F causa lo que sea que G cause. Así, por ejemplo, si la secuencia específica de golpes causa que el espía entre en pánico, entonces la información de que el mensajero ha llegado causó que el espía entre en pánico. Por otro lado, si el mero golpe en la puerta causa que el espía entre en pánico (si dos golpes rápidos, por ejemplo, hubiesen tenido el mismo efecto) entonces no es la información de que el mensajero ha llegado lo que causa el pánico. De modo semejante, si el patrón de golpes causa que el espía crea que el mensajero ha llegado, entonces su creencia ha sido causada por la información de que el espía ha llegado. Por otro lado, si la creencia no es causada por el patrón de golpes sino por el mero golpe (cualquier patrón hubiese producido la misma creencia) entonces si bien la creencia es producida por los golpes del mensajero, no es producida por la información de que este ha llegado.

Es claro que esta noción de conocimiento sigue una línea confiabilista. La propiedad G de una señal s que porta la información de que p no se instanciaría si no fuese el caso de que p . Por lo tanto, la creencia de S de que p será verdadera en todas las situaciones contrafácticas en las que es causada por G . El hecho de que G cause la creencia de S hace que este tenga conocimiento porque hace que esa creencia sea verdadera en varias situaciones actuales y contrafácticas. Es la condición de esta alta proporción de creencias verdaderas el corazón de toda propuesta confiabilista. Queda sin embargo la pregunta de qué variedad de confiabilismo constituye la propuesta de Dretske. En lo que sigue

argumentaré que este enfoque es semejante al confiabilismo de procesos y la epistemología de la virtud en los aspectos relevantes para las críticas que he propuesto en la sección anterior.

Dretske señala que a partir de su ejemplo podemos ver que la creencia C de S de que p producida por la información de que p no porta necesariamente ella misma la información de que p . Que C misma porte la información de que p depende de qué otras cosas, además de la información de que p , pueden causar C . Por ejemplo, supongamos que el agente sabe específicamente qué patrón de golpes en la puerta porta la información de que el mensajero ha llegado y que por lo tanto, cuando escucha este patrón, adquiere la creencia de que ha llegado. Supongamos además que un doble agente le ha proporcionado a este espía información falsa sobre una señal alternativa para determinar la presencia del mensajero. Se le ha dicho que además del patrón de golpes, el mensajero podría emitir la expresión “código azul”. Sin embargo, “código azul” es justamente la expresión que emitiría cualquiera de muchos asesinos que quiere acceder a nuestro espía para matarlo. Es decir, esta expresión no porta la información de que el mensajero está presente. Sin embargo, hubiese causado en el espía la misma creencia. Por lo tanto, la creencia del espía, aún cuando sea causada por la información de que el mensajero está presente, no porta ella misma esta información. Dretske consideraría que en una situación como esta, dado que la creencia es causada por la información relevante, el espía tendría conocimiento.

Ahora bien, es claro que si aceptamos el argumento que ofrecí en la sección anterior en contra de la suficiencia de la confiabilidad de las facultades cognitivas para el conocimiento, tenemos que aceptar que la condición de Dretske tampoco es suficiente. El caso de Sammy que vimos era un caso en el que un sujeto adquiere una creencia verdadera, la creencia de que los Yankees ganaron, a partir de un proceso confiable, el proceso de leer los resultados en el diario, pero que sin embargo no tiene conocimiento porque en la situación contrafáctica en la que lee que los Yankees pierde forma automáticamente la creencia de que ganaron, sobre la base de sus deseos. El sujeto no conoce porque a pesar de que su creencia no hubiese sido falsa si fuese formada por el mismo proceso que el de la situación actual, forma su creencia por medio de otro proceso en una situación contrafáctica en la que es falsa.

Pero el caso del espía recién descrito tiene las mismas características que el caso de Sammy. La creencia del espía es verdadera en la situación actual y en las situaciones contrafáticas en la que es producida por la misma señal. Sin embargo, hay muchas situaciones contrafáticas, aquellas en las que

los asesinos llegan antes que el mensajero, en la que la creencia es falsa y el espía la adquiere a partir de otra señal, una señal que no porta la información de que el mensajero está presente. Esto quiere decir que la creencia del sujeto no es segura, habría sido fácilmente falsa. El hecho de que la creencia del espía sea verdadera en la situación actual es sumamente accidental. Pero, como argumenta Pritchard, este tipo de suerte epistémica es incompatible con el conocimiento. Por lo tanto, la condición de Dretske es insuficiente.

La condición adicional que este ejemplo exige es claramente la condición de seguridad. La creencia del espía no constituye conocimiento porque hay muchas situaciones contrafácticas cercanas en las que es falsa. Vimos que lo que falta en este ejemplo es también que la creencia del espía porte la información de que el mensajero está presente. Esta condición es muy parecida a la de seguridad. Que la creencia no se instanciaría a menos que el hecho conocido ocurra, que hay una correlación entre el hecho y la creencia, *implica* que la creencia porta la información de que el hecho se da. La información no es otra cosa que el resultado de una correlación de este tipo. Sin embargo, estas dos condiciones no son equivalentes. Una creencia puede portar información sobre un hecho sin ser una creencia segura. Por ejemplo, supongamos que el espía sólo adquiriría la creencia de que el mensajero ha llegado a causa de la emisión de “código azul”. Como esta expresión sólo sería emitida por un asesino, entonces la creencia del espía de que hay un mensajero presente porta la información de que hay un asesino presente. Sin embargo, la creencia está lejos de ser segura, es falsa en toda situación contrafáctica cercana en la que es adquirida. Este desfasaje entre la información que porta la creencia y la seguridad se da porque la seguridad, como las diferentes nociones de confiabilidad, es entendida en términos de verdad, del ratio de creencias verdaderas en diferentes situaciones. Es decir, es entendida en términos de la coincidencia de un contenido con un hecho en diferentes situaciones actuales y posibles. Por el contrario, que una creencia porte cierta pieza de información es totalmente independiente del contenido de esa creencia. Sólo requiere que haya una correlación entre esa creencia y cierto hecho. Como claramente no diríamos que el espía en esta situación descrita conoce, la condición de que la creencia porte información es claramente insuficiente para el conocimiento. Lo que necesitamos para que haya conocimiento es justamente que la información que porta la creencia *coincida* con el contenido de la misma. Ahora bien, dado que he ofrecido una caracterización informacional del contenido, esta coincidencia no es algo difícil de explicar. Vimos que el contenido de una creencia es la información que tiene la función de portar (las piezas de información que causaron sus primeras instancias). Por lo tanto, el contenido de una creencia coincide con la información que efectivamente porta cuando esta

información es la misma que tiene la función de portar.

Ahora sí estamos en condiciones de ofrecer una condición informacional para el conocimiento que sea equivalente a la condición de seguridad:

Un sujeto S conoce el hecho de que p si tiene una creencia C que porta la información de que p y C tiene la función de portar la información de que p .

Vimos que el primero de los argumentos que ofrecí en contra de la epistemología de la virtud establece que (i) la confiabilidad de las facultades cognitivas no es suficiente para el conocimiento y que (ii) la condición de seguridad es necesaria. Por su parte, vimos que el segundo argumento que propuse muestra que (i) la confiabilidad de las facultades cognitivas no es necesaria para el conocimiento y que (ii) la seguridad es necesaria. Por medio de un argumento análogo al primero vimos en esta sección que la condición de Dretske, la condición de que la señal que causa la creencia porte información (en adelante “condición de información pre-doxástica”) es también insuficiente para el conocimiento y que seguridad es necesaria. Además, como vimos que la condición de seguridad y la condición de que la creencia porte la información que tiene la función de portar (en adelante “condición de información doxástica”) son equivalentes esta última es también necesaria para el conocimiento. El último argumento mostró entonces que la información pre-doxástica no es suficiente y que la información doxástica es necesaria. Surge entonces naturalmente la pregunta de si hay un análogo al segundo argumento que propuse en la sección anterior que muestre que la información doxástica es suficiente y que la pre-doxástica no es necesaria. T

Sin embargo, creo que no hay un argumento de este tipo puesto que la relación entre estas dos condiciones es más compleja. Considero que aunque la información doxástica es suficiente, la información pre-doxástica es necesaria. La razón es simplemente que aunque, como vimos, la información pre-doxástica puede darse sin información doxástica, la información doxástica implica información pre-doxástica, no son dos condiciones independientes. Una creencia C no puede portar información sobre el hecho de que p si la señal s que la causa no porta la misma información. Si s no porta información de que p entonces esto quiere decir que s puede ser causada por otros hechos además del hecho de que p . Por ejemplo, puede ser causada por el hecho de que q . Pero si la instanciación de s

causa la instanciación de C entonces C también puede ser causada por el hecho de que q y por lo tanto tampoco porta la información de que p .

Para ver este punto más claramente, veamos porque no podemos armar un ejemplo paralelo al ejemplo de Sam que ofrecí en la sección pasada, en contra de la propuesta de Dretske. El caso de Sam es el caso de un sujeto cuyas experiencias perceptuales y creencias perceptuales se forman de un modo no confiable. En presencia de un granero tiene la experiencia de que hay un elefante (y viceversa) y cuando tiene la experiencia de que hay un elefante adquiere la creencia de que hay un granero (y viceversa). Como vimos, el resultado de estos dos procesos no confiables es que Sam adquiere confiablemente creencias verdaderas sobre la presencia de graneros o elefantes. Ahora bien, aunque el proceso de formación de experiencias sobre graneros no sea confiable (pues la presencia de un granero produce sistemáticamente una experiencia que tiene un contenido falso o erróneo) las experiencias que producen los graneros sí portarán la información de que hay un granero. Las experiencias de como si hubiese un elefante son producidas sólo por graneros (en particular, no son producidas por elefantes). Por lo tanto, la creencia de que hay un granero sí es causada por un estado (la experiencia de como si hubiese un elefante) que porta la información de que hay un granero. Lo mismo ocurrirá con la creencia de que hay un elefante presente. Por lo tanto, el caso contra la necesidad de la confiabilidad de las facultades cognitivas no funciona contra la necesidad de la información pre-doxástica.

El desfazaje entre confiabilidad e información se da por la misma razón que mencioné antes. La confiabilidad se mide en relación a la verdad o corrección de los outputs en diferentes situaciones. Es decir, se mide en relación a la coincidencia entre el contenido de una representación y el hecho representado en diferentes situaciones. Por el contrario, la información sólo necesita una correlación entre dos eventos que es independiente de cualquier contenido. Sin embargo, en este caso, a diferencia del caso de la información doxástica, no parece necesario requerir una coincidencia entre la información pre-doxástica y el contenido del estado que la porta. En el caso de la información doxástica esta coincidencia es requerida porque el conocimiento requiere que la creencia sea verdadera y que no lo sea accidentalmente. Si esto puede ocurrir sin que los estados que causan la creencia sean verdaderos o correctos, como vimos en el caso de Sam, esto no constituye un problema.

Entonces, la información pre-doxástica es necesaria pero la información doxástica es suficiente porque esta implica a la pre-doxástica. No se trata de dos condiciones que sean necesarias y sólo *conjuntamente* suficientes para el conocimiento. No hay caso en el que se de la información doxástica y

no se de el conocimiento. Por lo tanto, la condición de información doxástica basta para caracterizar al conocimiento. El conocimiento es una creencia que porta la información que tiene la función de portar.

La noción de conocimiento integra dos dimensiones informacionales, la dimensión del contenido y la dimensión de la justificación. Sin embargo, hemos visto que la noción informacional de contenido conceptual que he propuesto es más compleja que la de Dretske. Por lo tanto, es de esperar que la noción informacional de conocimiento que integre esa noción de contenido sea algo más compleja que la que la que he presentado hasta aquí.

Como vimos, la información que tiene que portar una creencia B para ser conocimiento es la información que se corresponde con su contenido, es decir, con la información que tiene la función de portar. Ahora bien, argumenté en el capítulo 2 que el contenido de un concepto C que constituye B no está determinado sólo por la información que él mismo tiene la función de portar, sino también por la información que tienen la función de portar muchas de las estructuras con las que C está causalmente relacionado, estructuras que forman parte del rol funcional de C . En particular, dije que el contenido intencional de C está constituido por la información que él tiene la función de portar y el contenido cognitivo de C está constituido por la información que tienen la función de porta las estructuras que figuran en las *inferencias por default* de C . Las inferencias por default son las inferencias que figuran más frecuentemente en las interacciones con los miembros de la clase a la que C refiere, la clase que constituye su contenido intencional. De este modo para que una creencia, constituida por un concepto C constituya conocimiento, es necesario que la información que porte la creencia sea la información que todas estas estructuras, C y las estructuras en las inferencias por default de C , tienen la función de portar.

Es claro que no es posible que se instancien a la vez todas las inferencias por default de C , pues esto superaría la capacidad de la memoria de trabajo. Por lo tanto, la instanciación de C no puede requerir la instanciación de esta parte de su rol funcional. Pero para que una estructura porte información, tiene que instanciarse. Una estructura porta información cuando se activa o se instancia en presencia de una propiedad y no se hubiese activado si esa propiedad no hubiese estado presente. Esto quiere decir que para que una instancia de C constituya conocimiento no puede exigirse que las estructuras mismas de sus inferencias por default porten la información que tienen la función de portar. C mismo debe portar esta información. Por lo tanto, para que una creencia B constituya conocimiento, el concepto C que la constituye debe portar tanto la información que C mismo tiene la función de

portar, como la información que tienen la función de portar las estructuras en sus inferencias por default.

De este modo, el mapa informacional del conocimiento tiene la siguiente estructura. Por un lado, el conocimiento involucra dos primeras dimensiones informacionales, la dimensión del contenido, la de la función de portar información, y la dimensión de la justificación doxástica, la de la información efectiva. El conocimiento requiere que estas dimensiones coincidan. A su vez, la dimensión del contenido de un concepto *C* se sub-divide en dos, la información que *C* tiene la función de portar y la información que las estructuras en sus inferencias por default tienen la función de portar. Como el conocimiento requiere una coincidencia de contenido e información efectiva, la dimensión informacional de la justificación se sub-divide también en dos. *C* tiene que portar tanto la información que tiene la función de portar como la información que las estructuras de sus inferencias por default tienen la función de portar.

4. Tipos de conocimiento y un anti-empirismo epistémico

Dada la relación antes establecida, en el capítulo pasado, entre representaciones perceptuales y conceptuales, y la relación establecida, en la sección anterior, entre el conocimiento y lo conceptual, podemos preguntarnos ahora, respecto de la relación entre el conocimiento y lo perceptual ¿Qué hace a un caso dado de conocimiento, un caso de conocimiento *perceptual*? Típicamente distinguimos tipos de conocimiento por los tipos de procesos cognitivos o facultades cognitivas que lo producen. Por ejemplo, consideramos que el conocimiento perceptual es el producido por un proceso o sistema perceptual, el inferencial por un proceso o facultad inferencial, el testimonial es el que es producido por un proceso o facultad asociada al testimonio. Nótese que la idea no es meramente que la percepción, por ejemplo, produce las creencias que constituyen conocimiento perceptual, sino que produce el *conocimiento* perceptual mismo. Es decir, en los casos de conocimiento perceptual, la percepción no tiene sólo un papel causal sino, crucialmente, también *epistémico*, contribuye en algún sentido a que la creencia producida sea un caso de conocimiento.

De la caracterización propuesta en las secciones anteriores, se sigue que si por “conocimiento *perceptual*” entendemos un tipo de conocimiento en el que el proceso perceptual que produce la

creencia desempeña un papel epistémico, entonces no hay ningún caso de este tipo, como de ningún tipo específico de conocimiento. Pues el conocimiento es determinado sólo por el contenido informacional de un concepto, la información que efectivamente porta y la coincidencia de ambos. No hay un papel epistémico para el *proceso cognitivo* que produce dicha información. Lo que típicamente distingue a los tipos de conocimiento, los diferentes procesos cognitivos que producen el conocimiento, no entra en la individuación de casos de conocimiento.

Esto es simplemente una consecuencia del rechazo de las epistemologías de la virtud y del confiabilismo de procesos. En estas concepciones, individuar el proceso o facultad que produce la creencia *es epistémicamente relevante* para determinar la confiabilidad de la creencia resultante y por lo tanto, determinar que esa creencia es un caso de conocimiento. Según estos enfoques, para determinar si una creencia verdadera es conocimiento, o bien debemos evaluar el valor de verdad de esa creencia en las situaciones contrafácticas en las que es producida *por el mismo tipo de proceso*, o bien debemos evaluar el valor de verdad de cualquier creencia producida en una situación contrafáctica *por el mismo tipo de proceso*. Como vimos, esta concepción es vulnerable a la suerte epistémica. Debemos considerar sólo y todas las situaciones contrafácticas cercanas en las que es producida la misma creencia, *independientemente del proceso o facultad que la produzca en cada una de esas situaciones*. De este modo, el factor que típicamente empleamos para distinguir tipos de conocimiento queda afuera de la caracterización de la noción general de conocimiento, pierde su rol epistémico. Al menos en este sentido, ya no tenemos tipos de conocimiento y, en particular, no tenemos conocimiento perceptual.

Es plausible pensar que una formulación mínima de un empirismo epistémico afirma que todo conocimiento proviene, directa o indirectamente, de los sentidos. Pero, como vimos, cuando decimos que un conocimiento proviene de los sentidos no queremos decir meramente que la creencia que lo constituye fue causada por un sistema perceptual. Queremos decir que el que dicha creencia sea causada por un sistema perceptual es lo que la hace un caso de conocimiento. Pero como vimos, ningún sistema o proceso constituye parte de las condiciones que hacen que una creencia sea conocimiento. Por lo tanto, no sólo es el caso de que no todo conocimiento proviene de los sentidos, sino que ningún caso de conocimiento lo hace.

5. Conclusión

En este capítulo he intentado ofrecer una concepción informacional del conocimiento. Para ello, en primer lugar, he argumentado a favor del enfoque de la seguridad y en contra de las versiones confiabilistas de la epistemología de la virtud. He criticado los argumentos de Pritchard en este sentido y he propuesto dos argumentos alternativos que muestra, respectivamente, la necesidad y la suficiencia de la condición de seguridad para el conocimiento. En segundo lugar, he ofrecido una crítica al enfoque informacional de Dretske (1981) sobre la base del primer argumento a favor de la noción de seguridad. Considero que el enfoque de Dretske adolece de un problema semejante al de la epistemología de la virtud. Es vulnerable a la suerte epistémica. En tercer lugar, ofrecí una condición informacional para el conocimiento equivalente a la noción de seguridad. Como resultado, propuse una noción de conocimiento que incluye una caracterización informacional del contenido y de la justificación. Este último punto es crucial para reparar en el valor de esta propuesta informacional. Considero que lo que hace interesante la elucidación informacional de la noción de seguridad es justamente que proporciona una concepción integrada de diferentes dimensiones de los estados cognitivos de creencia. Dretske (1981) intentó, como vimos, llevar adelante una tarea semejante pero, como nos dejó ver la discusión sobre la noción de seguridad, este proyecto fue desarrollado de forma insatisfactoria.

Sin embargo, considero que el proporcionar una caracterización de diferentes dimensiones de los estados mentales en términos de la noción de información no es el único aspecto integrador de la propuesta. La propuesta informacional específica que he ofrecido aquí, a diferencia de la de Dretske, desvela una relación más estrecha entre el contenido y la justificación. En particular, muestra que la relación entre el contenido y la justificación de una creencia dada es la relación entre su función y el cumplimiento efectivo de dicha función. Vimos además que la noción informacional de contenido conceptual que he defendido en capítulos anteriores complejiza más esta relación entre contenido y justificación. Pues para estar justificada, una creencia dada no tiene que cumplir sólo con la función que ella tiene, sino con la función que tienen otras estructuras, las estructuras que constituyen su contenido.

Por último, he argumentado que de esta caracterización informacional del conocimiento se sigue que, dado que las facultades y procesos cognitivos son epistémicamente irrelevantes en el sentido de que no son necesarios para determinar que una creencia es un caso de conocimiento, entonces no tenemos un criterio adecuado para distinguir *tipos* de conocimiento. De esto se sigue, en particular, que no podemos decir que haya propiamente conocimiento perceptual. Es decir, se sigue una forma de anti-empirismo epistémico.

De todo esto resulta un complejo mapa de propiedades informacionales que constituyen nuestros estados cognitivos, con el que concluyo al menos una primera parte del enfoque informacional de la mente que quiero proponer. Queda mucho aún por decir sobre las relaciones entre los diferentes estados mentales así caracterizados y hay tipos de estados mentales que aún no han entrado en el foco de este proyecto informacional. En particular, queda aún por desarrollar una caracterización del tipo de estado más saliente en nuestra vida mental: las emociones. Considero que el enfoque informacional tiene mucho para decirnos sobre estos estados y que las elucidaciones llevadas a cabo en esta primera parte pueden ser de mucha ayuda para esta tarea. Sin embargo, por el momento, detendré en este punto la tarea de elucidar los estados que constituyen los diferentes niveles de procesamiento cognitivo y pasaré a considerar la relación de la noción de información con los diferentes niveles de realización de estos procesos. En particular, examinaré la relación de la noción de información con las de computación y procesamiento neuronal.

Parte II

Información y sus bases de realización

La relación entre cognición, computación y procesamiento neuronal

“En el principio era la tesis de la identidad-psicofísica” Sergio Barberis (2014)

“Dios es aquello a través de lo cual las cosas que de otro modo serían imaginarias son realizadas” Leibniz (1697/1989)

Como señalé en el primer capítulo, en esta segunda parte me ocuparé del vínculo entre la cognición, informacionalmente entendida y los tipos de procesos que se supone que constituyen su base de realización: los procesos computacionales y los procesos neuronales. Como vimos, determinar que la cognición, entendida informacionalmente, puede realizarse de este modo, es un elemento indispensable para adoptar la concepción informacional de la cognición. Pues, como vimos, la ventaja crucial del enfoque informacional, respecto de las propuestas alternativas, es la posibilidad de ofrecer una concepción naturalista del contenido mental, la posibilidad de compatibilizar una psicología intencional con una perspectiva naturalista.

En el capítulo 7 examinaré desarrollos recientes en torno a la noción de computación y su relación con el procesamiento neuronal. En particular, presentaré y adoptaré la caracterización que Piccinini y Bahar (2013) ofrecen de la noción de computación. Sin embargo, mostraré que la distinción entre información y computación es más fuerte que la que sostienen estos autores. No es sólo el caso de que la noción de computación no implica la de información sino que la de información tampoco implica computación. A su vez, mostraré que esta noción de computación implica una relación más estrecha entre procesamiento neuronal y computación que la que comúnmente se sostiene. El procesamiento neuronal no constituye una base de realización para la computación sino que es simplemente un tipo específico de computación, junto a la digital y la analógica.

En el capítulo 8 evaluaré la tesis fodoriana de que la cognición, informacionalmente caracterizada, no implica ni es implicada por la computación y la evaluaré, defenderé y refinaré a la luz de los desarrollos recientes discutidos en el capítulo 7. Considero que la noción fodoriana de computación, en términos de manipulación de formas sintácticas, debe ser abandonada a favor de la de Piccinini y Bahar, pero que, en contra de lo que implican tesis de estos autores, su noción no es implicada por la de cognición, informacionalmente entendida. Además, considero que si bien la noción

de forma sintáctica no nos sirve para caracterizar a la computación, sí nos sirve para caracterizar aspectos importantes de la noción informacional de cognición. La cognición sí implica la manipulación de formas sintácticas y esta característica pone restricciones relevantes para la tesis de que la cognición es un proceso computacional.

En el capítulo 9 argumentaré que aunque no hay una relación de implicación de computación a cognición, sí puede haberla de procesamiento neuronal a cognición, si entendemos al procesamiento neuronal no meramente en virtud de los vehículos que manipula sino también a partir de la función del sistema que lo lleva a cabo. El procesamiento neuronal es la actividad de un sistema, el sistema nervioso, cuya función peculiar es el control por retroalimentación negativa. Considero que el tipo específico de retroalimentación negativa que lleva a cabo el sistema nervioso implica cognición. Si aceptamos que esta función es un aspecto constitutivo del procesamiento neuronal entonces éste implica a la cognición. Antes de argumentar a favor de esta relación de implicación, sin embargo, defenderé una caracterización peculiar de la noción de retroalimentación negativa, en el contexto del debate que Wimsatt (1971) llevo adelante en torno a esta noción.

Por último, en el capítulo 10, me introduciré en el debate contemporáneo en torno a la noción de realización para determinar qué nos dicen las relaciones conceptuales establecidas en los tres capítulos anteriores sobre las relaciones metafísicas que mantienen la cognición, la computación y el procesamiento neuronal. Presentaré y discutiré los dos enfoques más difundidos actualmente: el enfoque plano y el dimensionado. En particular, argumentaré que si bien debemos abandonar el enfoque dimensionado, por ser incompatible con la tesis del fisicalismo de casos, el enfoque más adecuado carece de varios de los supuestos constitutivos del enfoque plano. De estas tesis resulta una especie de enfoque híbrido a partir de cual podemos determinar que las nociones relevantes estudiadas en los capítulos anteriores mantienen tres relaciones diferentes entre sí. La computación efectivamente no es realizada por el procesamiento neuronal sino que este es simplemente un tipo específico de aquella. La cognición, por su parte, sí se realiza en procesos computacionales y neuronales, pero de un modo diferente en cada uno de ellos. Se realiza de un modo metafísicamente necesario en el procesamiento neuronal (la instanciación del procesamiento neuronal es metafísicamente suficiente para la instanciación de un proceso cognitivo o informacional) y se realiza de un modo metafísicamente contingente sobre la computación (la instanciación de un proceso computacional es sólo nomológicamente suficiente para la instanciación de un proceso cognitivo o informacional). Estas dos últimas tesis eran justamente las que necesitamos para defender un enfoque informacional de la

cognición. Tenemos ahora un suelo firme, desde el punto de vista naturalista, para adoptar un enfoque informacional porque podemos explicar de qué modo este enfoque permite la naturalización del contenido, podemos explicar por qué los procesos informacionales pueden ser idénticos a procesos de niveles inferiores, físicos, neuronales o computacionales.

Capítulo 7

Computación y Procesamiento neuronal

1. Introducción

En las últimas décadas, el computacionalismo ha sido uno de los enfoques sobre los procesos cognitivos (y sobre su base neuronal) más ampliamente aceptado¹. El computacionalismo puede ser caracterizado principalmente por medio de dos tesis centrales. La primera es la idea de que el procesamiento neuronal es computacional. La segunda es que los procesos cognitivos son procesos computacionales. En filosofía, un tipo de contribución importante que se ha hecho al debate sobre estas tesis ha consistido en la elucidación de las nociones involucradas en ellas y la determinación de relaciones conceptuales que hubiese entre dichas nociones.

Considero que este tipo de tarea de elucidación conceptual tiene al menos dos funciones diferentes. La primera es la de contribuir a la argumentación a favor del computacionalismo. Veremos que en la literatura reciente en torno a las tesis computacionalistas, las relaciones conceptuales entre las nociones relevantes desempeñan un papel central en argumentos a favor de dichas tesis. Por ejemplo, Piccinini y Bahar (2013) ofrecen dos argumentos a favor de la primera tesis computacionalista. El primero parte de establecer una relación conceptual entre las nociones de vehículo neuronal y vehículo computacional y el segundo parte de establecer una relación entre las nociones de procesamiento de información y computación. En segundo lugar, el establecimiento de relaciones conceptuales entre las nociones relevantes puede ser importante para determinar el tipo de relación metafísica que guardan el

¹ Defensas clásicas e influyentes del computacionalismo incluyen a McCulloch y Pitts (1943), Wiener (1948), von Neumann (1958), Marr (1982), Churchland and Sejnowski (1992), Fodor (1975), Newell and Simon (1976), Pylyshyn (1984), and Newell (1990). Defensas más recientes del computacionalismo incluyen a Edelman (2008), Gallistel and King (2009), Schneider (2011) y, como veremos, Fodor (2008).

procesamiento neuronal, la computación y la cognición. Por ejemplo, Fodor (1995), habiendo caracterizado a la cognición en términos de procesamiento de información, mostró que la cognición así caracterizada guarda con la computación (tal como él la caracteriza) una relación confiable, pero metafísicamente contingente y que este tipo de relación es incompatible con la concepción estándar de la implementación.

Mi objetivo en este capítulo es examinar las nociones involucradas en la *primera* de las tesis del computacionalismo, la tesis de que el procesamiento neuronal es computacional y extraer algunas consecuencias que estas tienen para los dos argumentos mencionados de Piccinini y Bahar (2013) a favor de esta tesis.

2. Computación, procesamiento de información y procesamiento neuronal

Como señalé más arriba, Piccinini y Bahar (2013) ofrecen dos argumentos a favor del computacionalismo genérico, la tesis de que el procesamiento neuronal es computacional en el sentido más genérico de la noción de computación. Veamos entonces primero, en qué consiste la computación así entendida.

La noción más general de computación, la noción de computación genérica, es introducida por Piccinini y Scarantino (2011) como un modo de capturar “todos los usos relevantes de 'computación' en ciencias cognitivas” (2011, p. 10). Piccinini y Bahar (2013) caracterizan más específicamente las reglas que definen la computación. Afirman que las reglas (los mapas de input-output) que definen la computación son sensibles sólo a dimensiones específicas de variación de los vehículos y no a otras, *son insensibles a propiedades físicas más específicas de los vehículos* (pp. 6-7). Piccinini y Scarantino llaman a este tipo de vehículos “independientes del medio” porque esta propiedad característica hace posible la realizabilidad múltiple de los vehículos computacionales, permiten que una computación dada sea implementada por diferentes medios físicos (mecánicos, electromecánicos, electrónicos, magnéticos, etc.) . Un tipo de vehículo computacional puede ser realizado en una variedad de medios porque su realización no requiere de propiedades físicas muy específicas o concretas.. Piccinini y Bahar (2013) considera que esta es la noción más general de computación ya que señala lo que es común a los vehículos de los dos tipos más generales de computación: la computación digital y la

analógica. Veamos ahora en qué consisten estos tipos de computación y por qué ambos caen bajo el concepto de computación genérica.

Esquemáticamente, la computación digital abstracta es la manipulación de series de elementos discretos de un conjunto finito de tipos. Como Piccinini y Bahar están interesados en determinar qué tipo de computación, si alguno, puede ser realizada por el sistema nervioso, están interesados principalmente en la computación digital concreta o computación digital física. Estos elementos discretos pueden ser implementados por lo que llamamos “dígitos”. En una primera aproximación, la computación digital concreta es el procesamiento de secuencias de dígitos de acuerdo con reglas generales definidas sobre dígitos (Piccinini 2007). Un dígito, el vehículo atómico de la computación digital, es un estado macroscópico (un componente del sistema) cuyo tipo puede ser confiable y no ambiguamente distinguible por el sistema de otros tipos de estados macroscópicos. Un gran número de estados microscópicos se corresponden con cada tipo (macroscópico) de dígito. Por ejemplo, un número enorme de arreglos de electrones (estados microscópicos) se corresponden con la misma carga almacenada en un capacitor. Los sistemas digitales están contruidos de modo tal que responden del mismo modo a estos estados microscópicos diferentes (el que corresponde a su tipo de dígito). Como las propiedades más específicas (microscópicas) del medio no son relevantes para que este implemente un dígito Piccinini y Bahar sostienen que los dígitos son independientes del medio.

El otro tipo de computación que comúnmente se pone en contraste con la computación digital es la computación analógica. Piccinini y Bahar consideran que la noción más clara de computación analógica es la de Pour-El ((Pour-El 1974, Rubel 1993, Mills 2008). De acuerdo con esta noción, la computación analógica abstracta es la manipulación de variables continuas (variables que varían continuamente en el tiempo y pueden tomar cualquier valor real entre un intervalo dado) especificadas por ecuaciones diferenciales, de modo tal que se instancien relaciones funcionales apropiadas entre las variables.

Las variables continuas físicamente implementadas son muy diferentes a los dígitos. Un sistema de computación digital siempre puede distinguir de modo no ambiguo un dígito de otro. Por el contrario, un sistema de computación analógica no puede hacer lo mismo con los valores exactos de las variables continuas, ya que estas variables sólo pueden ser medidas con un margen de error. Como las variables pueden tomar cualquier valor real pero hay un límite a la capacidad de discriminación de cualquier sistema, es siempre posible que la diferencia entre dos porciones de una variable continua sea

suficientemente pequeña como para no ser detectada por el sistema. Sin embargo, la computación analógica sólo es sensible a diferencias entre porciones de las variables continuas, al grado que le sea posible distinguirlas. Cualquier propiedad más específicas del medio que implementa a las variables es irrelevante para la computación. Por este motivo, Piccinini y Bahar considera que la computación analógica también es independiente del medio. Así la noción de computación genérica parece rescatar las dos nociones más generales de computación.

Ahora bien, como señalan Piccinini y Scarantino (2011) señalan que, en las últimas décadas, la analogía entre los cerebros y las computadoras se ha impuesto en neurociencia. Muchos neurocientíficos han empezado a usar el término computación para referirse al procesamiento de secuencias de potenciales de acción neuronales, secuencias de potenciales de acción producidos por las neuronas en tiempo real. El procesamiento de secuencias de potenciales de acción por el sistema nervioso es comúnmente llamado “computación neuronal”.

A partir de las consideraciones hechas sobre la noción de computación podemos entonces preguntarnos si es el procesamiento neuronal computacional en el sentido genérico. Piccinini y Bahar (2013) ofrecen dos argumentos a favor de una respuesta afirmativa.

El primer argumento es llamado “el argumento a partir de la organización funcional del sistema nervioso”. La idea es que ya que (i) los procesos neuronales son definidos sobre vehículos independientes del medio y (ii) el procesamiento de vehículos independientes del medio constituye computación en el sentido genérico, se sigue que (iii) el procesamiento neuronal es computación en el sentido genérico. La premisa (ii) es el resultado de la elucidación de Piccinini y Scarantino de la noción de computación genérica. Así, Piccinini y Bahar, necesitan sólo argumentar a favor de la premisa (i). Lo hacen señalando que evidencia actual indica que los vehículos primarias que manipulan los procesos neuronales son los potenciales de acción de las neuronas y que las propiedades funcionalmente relevantes de los potenciales son aspectos dinámicos de los mismos, las tasas y duración de los potenciales. Sólo estas dimensiones de los potenciales de acción y no propiedades más específicas, son relevantes para el procesamiento neuronal. Esto quiere decir, según Piccinini y Bahar, que los potenciales de acción son independientes del medio y por lo tanto el procesamiento neuronal es computacional en sentido genérico.

El segundo argumento de Piccinini y Bahar (2013) es llamado “el argumento desde el procesamiento de información”. Piccinini y Bahar afirman que como (i) los procesos neuronales

procesan información semántica y (ii) el procesamiento de información semántica requiere computación genérica, entonces se sigue que (iii) los procesos neuronales son computación en el sentido genérico. Aquí las dos premisas requieren argumentación.

En relación a la premisa (ii), Piccinini y Bahar afirman que la noción de información, en cualquiera de tres sentidos diferentes que puede dársele, implica independencia del medio. Piccinini y Scarantino (2011) distinguen tres nociones principales de información que Piccinini y Bahar (2013) toman: una noción no semántica, una noción semántica natural y una noción semántica no natural. La noción formalizada por Shannon (1948) es no semántica ya que una pieza de información en este sentido es plenamente caracterizada por dos características: es una estructura física distinguible de un conjunto de estructuras físicas alternativas y pertenece a un conjunto exhaustivo de estructuras físicas mutuamente exclusivas con probabilidades bien definidas. Ninguna propiedad semántica está involucrada en su individuación. Sin embargo, un teórico de la información podría estar interesado también en el contenido específico de una pieza de información. Piccinini y Scarantino (2011) toman la distinción de Grice (1957) entre significado natural y no natural y lo aplican para distinguir dos tipos de información semántica. Un caso de significado natural puede ser ejemplificado por la oración “esos puntos significan rubeóla” la cual es cierto sólo si el paciente tiene rubeóla. El significado no natural es ejemplificado por la oración “esos tres campanazos significan que el colectivo está lleno”, la cual es cierta incluso si el colectivo no está lleno. De modo semejante, podemos clasificar a la información como natural cuando hay una correlación física confiable entre dos estados. Los puntos portan la información sobre la rubeóla sólo cuando están correlacionados confiablemente con dicha enfermedad. Por el contrario, los campanazos portan información sobre el hecho de que el colectivo está lleno en virtud de una convención.

Según Piccinini y Bahar, procesar información, en cualquiera de estos sentidos, significa procesar vehículos en base a la información que portan y no sus propiedades físicas específicas. La información que porta una señal es definida independientemente del medio físico específico que implementa esa señal. Así, el procesamiento de información requiere procesar vehículos que son definidos de un modo independiente del medio. Por lo tanto, el procesamiento de información implica computación genérica (la inversa no vale porque los vehículos computacionales no necesitan portar información) (p. 12).

En relación a la premisa (i) Piccinini y Bahar mantienen que el procesamiento neuronal

involucra procesamiento de información en los tres sentidos mencionados antes. En primer lugar, la información mutua, la medida de dependencia estadística entre una fuente y un receptor en la teoría de Shannon (Shannon y Weaver 1949) es frecuentemente usada para cuantificar la dependencia estadística entre las señales neuronales y sus fuentes y estimar la eficiencia de codificación de las señales neuronales (Dayan y Abbott 2001, ch. 4; Baddeley et al. 2000). Hay también evidencia que muestra que las señales neuronales portan información semántica natural. Está bien establecido experimentalmente que las variables neuronales portan información semántica sobre otras variables internas y externas (Adrian 1928; Rieke et al. 1999; cf. Garson 2003). Mucho de la neurofisiología contemporánea se centra en el descubrimiento y manipulación de las variables neuronales que se correlacionan confiablemente con variables externas al sistema nervioso. Por último, Piccinini y Bahar afirman que hay muchas teorías en ciencias cognitivas que asumen que el sistema nervioso posee y procesa representaciones. En su interpretación de la noción de representación, esto quiere decir que el sistema nervioso posee y procesa información semántica no-natural. Específicamente, afirman que es difícil explicar el procesamiento del lenguaje y otros procesos cognitivos superiores sin postular representaciones. Como los potenciales de acción neuronales portan información en los tres sentidos principales y los vehículos del procesamiento neuronal son independientes del medio, Piccinini y Bahar concluyen que el sistema nervioso es computacional en el sentido genérico.

En la sección siguiente quisiera revisar críticamente la noción de computación genérica y su relación con la noción de información. Sobre la base de estas consideraciones, volveremos a considerar estos dos argumentos.

3. Independencia del medio y relevancia funcional restringida

Quisiera comenzar a considerar dos aspectos de la noción de información. Primero evaluaré la distinción de tipos de información que proponen Piccinini y Bahar, y luego la caracterización general que ofrecen de la noción de información. Como vimos, Piccinini y Bahar distinguen dos tipos de información semántica. La información semántica natural y la no natural. La natural es la que porta un estado a partir de cierta correlación con una propiedad u objeto dado. La no natural es que un estado adquiere por medio de una convención. La diferencia relevante, según los autores, es que sólo la información no natural puede ser falsa. Sin embargo, como vimos en el capítulo pasado, aunque es

cierto que una señal no puede portar información sobre un hecho sin que ese hecho ocurra (pues tiene que haber una correlación causal entre ambos), se han ofrecido y defendido diversas nociones de contenido informacional en términos de la noción natural de información. Podemos decir, por ejemplo, que el tipo de estímulo al que respondió por primera vez una estructura neuronal, constituye su contenido informacional y por lo tanto, cuando se activa en ausencia de ese estímulo tiene el contenido informacional falso de que ese estímulo está presente. Aunque noción de contenido ha enfrentado diversos problemas, considero que, a diferencia de la que ofrece Piccinini, es una noción informacional. La noción de información semántica de Dretske se vincula conceptualmente con la de Shannon ya que explota las correlaciones causales (y/o probabilísticas) entre una fuente y un receptor. La noción de *contenido* semántico informacional de Dretske se vincula con su noción semántica de información ya que la presupone (el contenido informacional es el tipo de información que portó por primera vez una estructura, la información natural que tiene la función de portar). Por otro lado, la noción de información no natural de Piccinini no tiene ningún vínculo claro con alguna de las nociones de información. El estado que porta información no natural puede no encontrarse en ningún tipo de correlación causal con el evento sobre el que porta información y puede no haber estado correlacionado causalmente con instancias previas de el mismo tipo de evento. Sólo basta que una convención los vincule. Por este motivo no nos parece adecuado sustituir, en la distinción de tipos de información, al contenido representacional dretskeano, basado en la información natural, por un tipo de “información” no natural, que no guarda vínculo conceptual con la noción de información de Shannon. Aunque este punto no afecta la argumentación que sigue, excluiré a la información no natural de Piccinini y Bahar del conjunto de tipos de información. Así, podemos hablar de una información semántica natural y de una información representacional natural.

Este punto es importante para los usos de la nociones de información y representación en las ciencias cognitivas. Piccinini y Scarantino (2011) admiten que los psicólogos y neurólogos cognitivos identifican frecuentemente la información natural con las representaciones, con items que pueden ser falsos, pero consideran que este es un uso erróneo de la noción de representación. La representación tiene que ser identificada, según ellos, con la información no natural. Sin embargo, considero que esta identificación, por parte de psicólogos y neurologos. puede estar fundada en la caracterización de la noción de representación en términos de información natural, en la noción dretskeana de información representacional.

Hechas estas consideraciones sobre los tipos de información, podemos preguntarnos ahora, en

general, qué determina si un objeto o evento e es un caso de cierto tipo de vehículo informacional I . La respuesta obvia es que el que e sea un caso de I depende del tipo de información que e porte. Pero la información que un objeto porta, como hemos visto, es plenamente determinada por ciertos poderes causales de e : O bien por tener la correlación causal confiable que e tenga un caso de un tipo de evento u objeto T (información semántica natural) o bien por haber tenido e originalmente una correlación causal confiable con otros casos de T (información representacional natural). Así, podemos decir que e pertenece a I si o bien guarda una correlación confiable un caso de T o bien guardó originalmente una correlación confiable con casos de T . ¿Hay algo más que sea necesario para que e pertenezca a I ? En particular, ¿Desempeñan algún papel las propiedades de e que son funcionalmente relevantes para los poderes causales que lo constituyen como un caso de I ?

Considero que este no es el caso. *Dos eventos a y b que comparten un poder causal c que los hace a ambos casos de un tipo informacional I , pueden no compartir ninguna propiedad funcionalmente relevante para c . Por lo tanto las propiedades funcionalmente relevantes para c , a diferencia de c mismo, no son constitutivas del tipo informacional I . Por ejemplo, el patrón de movimiento molecular, el sonido s , y el patrón de ondas de luz l no comparten ninguna propiedad funcionalmente relevante para la relación causal confiable que ambos tienen con una presa p (y hace que s y l porten información sobre la presencia de p). Otro ejemplo ordinario de esta característica de la información natural podría ser el siguiente: el humo y mi concepto FUEGO no comparten ninguna propiedad intrínseca relevante para la correlación confiable que ambos tienen con el fuego. El hecho de que la información no natural tiene también esta característica es más obvio, dado que las convenciones pueden ser más arbitrarias que las correlaciones naturales.*

Lo que quiero resaltar sobre este fenómeno (que vehículos que portan la misma información y pertenecen al mismo tipo informacional pueden no compartir propiedades funcionalmente relevantes) es que nos muestra que los vehículos informacionales son múltiplemente realizables. Los vehículos del procesamiento informacional pueden ser realizados en base físicas completamente diferentes, es decir, bases físicas que no comparten ninguna propiedad funcionalmente relevante para el poder causal que los constituye. Creo que esta característica es importante para comparar a los vehículos del procesamiento de información con los vehículos de la computación.

Concuerdo con Piccinini y Bahar en que, como afirman en su segundo argumento, hay una propiedad que los vehículos informacionales y computacionales comparten pero considero que esta es

insuficiente para que haya una relación de implicación de información a computación. Hemos visto que un tipo de vehículo informacional es constituido por un poder causal determinado y no por las propiedades funcionalmente relevantes para ese poder causal. Por el contrario, los vehículos de la computación sí son definidos en función de características de sus propiedades funcionalmente relevantes. Como vimos los vehículos de la computación son aquellos cuyas propiedades funcionalmente relevantes están restringidas a dimensiones específicas de variación de ciertas variables suyas.

Pero entonces, como no se requiere, para la instanciación de un poder causal c que constituye un tipo informacional I , la instanciación de ninguna propiedad específica funcionalmente relevante para c , no se requiere, en particular, que las propiedades funcionalmente relevantes para c estén restringidas a dimensiones de variación específicas de alguna variable suya. Es decir, no hace falta que una instancia de vehículo informacional sea computacional. Por lo tanto, información no implica computación.

Vimos que el hecho de que las propiedades funcionalmente relevantes de los vehículos computacionales estén restringidas a ciertos aspectos específicos, permite que estos sean múltiplemente realizables. Por este motivo, los vehículos computacionales pueden realizarse en bases físicas que sean diferentes sólo en aspectos *que no son funcionalmente relevantes*. Por el contrario, vimos que los estados informacionales pueden realizarse en diferentes bases físicas justamente porque *las propiedades que sí son funcionalmente relevantes* pueden variar totalmente de un caso a otro de un mismo tipo informacional. Así, si bien ambos tipos de propiedades son múltiplemente realizables, el tipo de realizabilidad hace que los tipos informacionales pueden no ser computacionales. Piccinini y Bahar afirman que lo que los vehículos de la información y la computación comparte es lo que llaman “independencia del medio”. Si por esto entendemos realizabilidad múltiple en general, la tesis es verdadera pero no habilita una implicación de información a computación. Si por “independencia del medio” queremos referirnos a la propiedad más específica que define a los vehículos computacionales, la de que sus propiedades funcionalmente relevantes estén restringidas de cierto modo, la tesis es falsa, ya que los vehículos informacionales pueden no estar restringidos de este modo. Considero que la expresión “independencia del medio” genera un equívoco entre estas dos lecturas. Por este motivo, me referiré a la característica que define a los vehículos de la computación como relevancia funcional restringida. Con esto sólo quiero decir que las propiedades funcionalmente relevantes de los vehículos computacionales están restringidas de cierto modo (a ciertas dimensiones de variación de una variable

específica suya).

Hechas estas consideraciones, podemos pasar ahora a revisar los argumentos de Piccinini y Bahar. Recordemos que el segundo argumento afirma que como (i) los procesos neuronales procesan información semántica y (ii) el procesamiento de información semántica requiere computación genérica, entonces se sigue que (iii) los procesos neuronales son computación en el sentido genérico. Aquí las dos premisas requieren argumentación.

La premisa (ii), que el procesamiento de información requiere computación genérica es más compleja de lo que puede parecer. Involucra tres tesis diferentes: (a) que los vehículos del procesamiento de la información semántica son independientes del medio que (b) los vehículos de la computación genérica son independientes del medio y que (c) esta característica común hace que los vehículos informacionales sean computacionales. Como hemos visto, cuando reconocemos que la independencia del medio es relevancia funcional restringida, Piccinini y Bahar están en lo cierto respecto de (b) pero no de (a). Los vehículos informacionales no necesitan tener relevancia funcional restringida. Sin embargo, mostré que estos dos tipos de vehículo sí comparten realizabilidad múltiple. Hay una característica compartida. Pero, como argumenté, en contra de (c), esta característica compartida no hace a los vehículos informacionales computacionales. No hay implicación de procesamiento de información a computación genérica ya que la independencia de medio de los vehículos del procesamiento de información es compatible con la carencia de relevancia funcional restringida. Por lo tanto este argumento no establece la tesis de que el procesamiento neuronal es computacional en el sentido genérico.

Considero sin embargo, que el primer argumento sí establece satisfactoriamente esta tesis. Recordemos que el primer argumento afirma que como (i) los procesos neuronales son definidos sobre vehículos independientes del medio y (ii) el procesamiento de vehículos independientes del medio constituye computación en el sentido genérico, se sigue que (iii) el procesamiento neuronal es computación en el sentido genérico. Ahora bien, creo que si bien este argumento es adecuado, establece una forma peculiar de la tesis computacionalista.

4. La idea de una base neuronal para la computación cognitiva

Como vimos en el capítulo 3, Fodor (2008) señala que es un lugar común en ciencias cognitivas considerar que los procesos computacionales que se realizan múltiplemente en el cerebro, pueden realizarse en diferentes tipos de estructuras neuronales. Ahora bien, considero que el primer argumento de Piccinini y Bahar (2013), hace algo más que establecer la tesis de que el procesamiento neuronal es computacional. Determina una modo peculiar en que la realizabilidad múltiple relaciona al procesamiento neuronal y la computación. Veamos por qué.

Como vimos en el capítulo 3, Fodor (2008) señala que es hoy en día generalmente asumido que, así como los procesos cognitivos pueden realizarse sobre diferentes estructuras de nivel inferior (neuronales o computacionales), los procesos computacionales que sea realizan en el cerebro pueden realizarse sobre diferentes bases neuronales. Considero que el argumento de Piccinini y Bahar tiene una consecuencia interesante en este respecto. Su argumento, como vimos, determina que el procesamiento neuronal es computacional porque procesa el mismo tipo de vehículos que los procesos computacionales, vehículos independientes del medio, o vehículos que tienen lo que llamé “relevancia funcional restringida”.

Lo que considero central de este argumento es que muestra que la noción de computación se aplica a la de procesamiento neuronal en el mismo sentido en que se aplica a los otros tipos de computación, digital y analógica. Los vehículos de la computación digital y analógica, al igual que los vehículos del procesamiento neuronal, son computacionales porque muchos aspectos de ellos no son funcionalmente relevantes. De este modo, a menos que digamos que la computación digital o analógica realizan a la computación genérica, no podemos decir que el procesamiento neuronal realice a la computación genérica. Sin embargo, resulta razonable decir que la computación digital y la analógica son simplemente tipos y no realizadores de la computación genérica. Por lo tanto, debemos decir lo mismo del procesamiento neuronal. Ahora bien, para fundar sólidamente la idea de que la computación digital, analógica y neuronal son tipos y no realizadores de la computación genérica, tendremos que esperar al capítulo 9. Allí veremos que esta idea se sigue también de la elucidación que propondremos de la noción de realización.

Del argumento de Piccinini y Bahar se sigue, entonces, que la relación entre lo neuronal y lo computacional no es la relación entre dos tipos de propiedades que se encuentran en diferentes niveles

sino que es la relación entre un género de propiedades y un tipo específico dentro de ese género. Esto contrasta abiertamente con lo que se cree comúnmente sobre la relación entre lo neuronal y lo computacional. Sin embargo, a pesar de ser controvertida, es una consecuencia conceptual bien fundada en la práctica científica misma. La noción de computación que ofrece Piccinini está basada efectivamente en el rasgo común de los dos tipos más generales de procesamiento que estudian las ciencias de la computación. La noción de vehículo del procesamiento neuronal está basada en evidencia empírica respecto de cuáles son los aspectos funcionalmente relevantes de los potenciales de acción neuronales. Si de estos dos elementos se sigue que el procesamiento neuronal está en el nivel computacional, parece que no tenemos otra alternativa más adecuada a la práctica científica que rechazar la distinción entre el nivel neuronal y el computacional.

Ahora bien, este no es la única consecuencia interesante que podemos extraer del argumento de Piccinini y Bahar. Es notable que, a pesar de lo dicho, la noción de realizabilidad no queda afuera de la relación entre lo computacional y lo neuronal. Aunque tengamos que aceptar que lo neuronal no se realiza en lo computacional, la razón por la que lo neuronal es computacional es justamente porque posee el tipo de realizabilidad múltiple propia de lo computacional, es independiente del medio. La realizabilidad múltiple de la computación cognitiva, entonces, no se da ella misma sobre diferentes bases de nivel neuronal sino que es de nivel neuronal y se da sobre una base de nivel inferior. A su vez, es esta misma realizabilidad la que hace que lo neuronal sea de nivel computacional. De este modo, si bien lo neuronal no se encuentra en una relación de realización con lo computacional sí están vinculados por medio de la noción de realización. Es un tipo de realizabilidad múltiple el criterio para determinar si un estado es de tipo computacional.

5. Conclusión

En este primer capítulo he comenzado el análisis de los diferentes niveles de implementación de lo cognitivo, desde una perspectiva informacional. He empezado por la elucidación de las nociones características de los dos niveles inferiores al cognitivo: el computacional y el neuronal. En primer lugar, he evaluado una reciente y, a mi parecer, prometedora propuesta sobre la noción de computación. Piccinini y Bahar ofrecen una noción anclada en la práctica de las ciencias de la computación, una noción que recoge el rasgo común de los dos tipos más generales de computación. Consideré que si

bien la noción es adecuada, su caracterización adolece de cierta falla. En particular, se confunde el tipo de realizabilidad múltiple que poseen los vehículos computacionales y los informacionales. Por este motivo, ofrecí algunos argumentos para resaltar la diferencia entre estos dos tipos de realizabilidad múltiple. La consecuencia de esta tarea no es sólo clarificar la noción de computación sino mostrar que uno de los argumentos (el segundo) a favor de la tesis de que los procesos neuronales son computacionales no puede funcionar, pues depende de la premisa que procesamiento de información implica computación. Sin embargo, considero que el primer argumento sí funciona para establecer esta tesis.

En segundo lugar extraje algunas consecuencias importantes sobre la relación entre lo neuronal y lo computacional que considero que se siguen de este primer argumento. Por un lado, mostré que se sigue que el nivel neuronal no es realmente diferente del nivel computacional. Los procesos computacionales no se realizan en procesos neuronales sino que los procesos neuronales son ellos mismos tipos de procesos computacionales. Aunque esta consecuencia se contradice con lo que se cree generalmente sobre la relación entre lo neuronal y lo computacional, es una consecuencia de la caracterización que de estos procesos ofrecen, respectivamente, la neurología y las ciencias de la computación. Por otro lado, mostré que a pesar de que lo computacional no se realiza en lo neuronal, estos dos tipos de propiedades sí están relacionadas por medio de la noción de realización. Es un tipo de realizabilidad múltiple lo que hace que lo neuronal se un tipo de proceso computacional.

En el próximo capítulo pasaremos ya a considerar la noción informacional de cognición y su relación con la noción de computación. Para este propósito serán cruciales las reflexiones aquí desarrolladas sobre la noción de computación y su relación con la de información.

Capítulo 8

Cognición y computación

1. Introducción

Como vimos en el capítulo pasado, hay tres puntos centrales a tener en cuenta respecto de la consideración filosófica de la tesis computacionalista. En primer lugar, el computacionalismo puede ser caracterizado principalmente por medio de dos tesis centrales. La primera es la idea de que el procesamiento neuronal es computacional. La segunda es que los procesos cognitivos son procesos computacionales. En segundo lugar, vimos que un tipo de contribución típicamente filosófica (aunque no el único) es la elucidación de las nociones involucradas en cada una de estas tesis y la determinación de relaciones conceptuales que hubiese entre dichas nociones. En tercer lugar, vimos que esta tarea de elucidación puede ponerse al servicio de al menos dos funciones diferentes. La primera es la de contribuir a la argumentación a favor del computacionalismo. Vimos en el capítulo pasado que, por ejemplo, Piccinini y Bahar (2013) ofrecen dos argumentos a favor de la primera tesis computacionalista basados en las relaciones conceptuales entre nociones involucradas. La segunda función que puede tener la elucidación conceptual es determinar el tipo de relación metafísica que guardan el procesamiento neuronal, la computación y la cognición.

En el capítulo pasado el foco de la tarea de elucidación y su uso estuvo en la primer tesis computacionalista. Distinguí la noción de computación de la de información, y mostré en base a ello que uno de los argumentos a favor de la primer tesis no funciona. Además, en base a la noción de computación de Piccinini y Bahar mostré el modo peculiar en que la noción de realización media la relación entre computación y procesamiento neuronal. En este capítulo pasaremos a considerar la segunda, la tesis de que la cognición es computación neuronal.

Respecto de la relación conceptual entre cognición y computación, hay una tesis que se sigue

inmediatamente a partir de tesis defendidas en capítulos anteriores. Por un lado, he defendido en los capítulos 2 y 3 una propuesta informacional de la cognición, entendida como procesamiento de representaciones conceptuales. Si unimos esto a la tesis defendida en el capítulo pasado de que la noción de información no implica la de computación (ni viceversa), se sigue que la cognición no implica computación (ni viceversa).

Ahora bien, hay un defensor ilustre de la tesis de que no hay una relación conceptual o metafísicamente necesaria entre estados computacionales y las representaciones que se realizan en ellos, entendidas informacionalmente. Como vimos, Fodor (por ejemplo 1994, 1998, 2008) ofrece buenas razones para identificar a la cognición con el procesamiento de información. A su vez, Fodor (1995) ha argumentado justamente que la instanciación de un estado informacional no implica la instanciación de un estado computacional (ni viceversa). Como veremos más adelante, uno de los problemas centrales de la argumentación fodoriana es que su noción de computación es trivial, se aplica prácticamente a cualquier sistema físico. Por el contrario, la noción de Piccinini y Bahar (2013) que vimos el capítulo pasado es extensionalmente más adecuada. Sabemos que, sin embargo, Piccinini y Bahar creen que su noción de computación, a diferencia de la de Fodor, sí es implicada por la noción de procesamiento de información. Argumenté el capítulo pasado que esta relación de implicación no se da tampoco para la noción de Piccinini y Bahar. De esto se sigue que, a diferencia de lo que sucede con la primera tesis computacionalista, no hay una relación conceptual entre las nociones involucradas en la segunda tesis computacionalista. Cognición no implica computación ni viceversa.

Esta tesis se sigue casi inmediatamente de lo dicho en capítulos anteriores. Sin embargo, considero que podemos ser aún más precisos en la caracterización de las posibles relaciones conceptuales entre información o cognición y computación. Distinguiré diferentes sentidos de la tesis de que información implica computación y argumentaré que la tesis es falsa según cada una de esas lecturas y según tanto la noción fodoriana de computación como según la de Piccinini y Bahar.

Creo, sin embargo, que esta ausencia de vínculo conceptual no es lo único que podemos decir sobre la relación entre cognición y computación. Considero que una comprensión más acabada de la cognición permite ver una restricción importante para que un proceso computacional implemente uno cognitivo. Esta comprensión más completa de la cognición involucra la noción fodoriana de forma sintáctica. Creo que si bien la noción fodoriana de forma sintáctica no es adecuada para caracterizar a la noción de computación, sí constituye un elemento crucial para caracterizar a la cognición, entendida

informacionalmente. Considero que procesamiento de información implica procesamiento de formas sintácticas. Así como en el capítulo pasado mostré que la computación cognitiva pertenece al nivel neuronal, en este capítulo mostraré que las formas sintácticas no pertenecen al nivel computacional, como sostiene Fodor, sino al nivel cognitivo/informacional. Ahora bien ¿En qué sentido esta caracterización más completa de la cognición es relevante para determinar cómo tiene que ser su base computacional? Mostraré que esta noción de forma sintáctica nos permite ver que para que la cognición (es decir *el procesamiento de información* y no meramente un estado informacional aislado) sea implementado por una base computacional debe haber una correlación confiable entre tipos informacionales y tipos computacionales.

En la sección 2 desarrollaré con algo de detalle la tesis fodoriana sobre la relación contingente entre computación y cognición, y la argumentación a favor de ella. En la sección 3 presentaré diferentes modos de entender dicha tesis, diferentes modos de entender la noción fodoriana de computación y mostraré cómo afectan a la argumentación fodoriana estas diferentes lecturas. En la sección 4 explicaré por qué la noción de Piccinini y Bahar es aún más adecuada que la lectura más caritativa de la noción fodoriana de computación pero argumentaré que aún si adoptamos esta noción, se sigue que la relación entre computación y cognición es contingente, según las diferentes lecturas que propuse en la sección 3. Por último, en la sección 5, desarrollaré la defensa de la tesis de que, a pesar de que la forma sintáctica no es relevante para caracterizar a la computación sí es crucial para caracterizar a la cognición. Al estar estrechamente vinculada a la noción de cognición, la noción de sintaxis se vuelve relevante para evaluar la relación entre cognición y computación. Mostraré que la noción de sintaxis determina un criterio relevante para evaluar la tesis de que la cognición es computación.

2. Fodor sobre computación y cognición

Antes de presentar la posición fodoriana debo hacer una aclaración. Dije previamente que la segunda tesis computacionalista es que los procesos cognitivos son computacionales. Sin embargo, como veremos, Fodor sostiene una tesis computacionalista más fuerte: la de que las leyes psicológicas son implementadas por procesos computacionales. Esta tesis contiene dos elementos que no menciono en mi formulación de la tesis: las nociones de ley y de implementación. En este trabajo quiero

permanecer neutral respecto de si la relación entre lo computacional y lo cognitivo es una de implementación y de si los procesos cognitivos son mejor caracterizados en términos de las leyes que los rigen. Pasaremos a tratar detalladamente estas cuestiones en el capítulo 9. Por el momento, adopto una formulación mínima de la tesis computacionalista, la idea de que los procesos cognitivos son computacionales. El objetivo de este trabajo es determinar si hay una relación conceptual entre las dos nociones involucradas en esta tesis y determinar las consecuencias que esta relación tiene para la tesis y para la relación metafísica entre computación y cognición. Para lograr este objetivo hay tres aspectos del enfoque fodoriano que sí adoptaré: (i) La caracterización intencional de los procesos cognitivos, (ii) la teoría informacional de la intencionalidad y (iii) la idea de que la relación entre los estados informacionales y los vehículos computacionales es metafísicamente contingente. Considero que estas tres tesis hacen plausible la idea de que la noción de cognición no implica la de computación. El problema, como veremos en la próxima sección, es que la noción fodoriana de computación es trivial y por lo tanto la tesis tendrá que ser revisada a la luz de una noción no trivial. Veamos ahora con un poco más de detalle la posición fodoriana.

Fodor (1994) caracteriza y defiende tres tesis diferentes sobre la explicación psicológica que parecen estar en conflicto entre sí. La primera es la idea de que (iF) la explicación psicológica está constituida, al menos en parte, por leyes intencionales. Como veremos la tesis (i) antes mencionada, la idea de que los procesos cognitivos son intencionales, es *parte* de (iF). La segunda tesis fodoriana es que (iiF) el contenido intencional puede ser caracterizado en términos de información. Esta tesis es equivalente a la tesis (ii) antes mencionada. La tercera es la tesis de que (iiiF) las leyes intencionales son implementadas por mecanismos computacionales. Esta tesis está relacionada de un modo diferente con la tesis (iii). La tesis (iii), la de que los estados informacionales y computacionales se relacionan contingentemente es justamente la que, aparentemente, está en conflicto con (iiiF) la idea de que los procesos cognitivos son implementados por procesos computacionales. Fodor considera que la adopción conjunta de iF, iiF y iiiF, junto con (iii), implican que los estados cognitivos y los computacionales se relacionan contingentemente y que esta tesis implica la negación de alguna o más de una de las tesis iF, iiF o iiiF. Según la noción estándar de implementación, para que las leyes intencionales sean implementadas por mecanismos computacionales, la instanciación de un estado intencional debe implicar necesariamente la instanciación de un vehículo computacional y viceversa. Pero como caracterizamos informacionalmente a los estados intencionales y como, según Fodor, no es metafísicamente necesario que dada la instanciación de un caso de determinado tipo de estado

informativa, se instancie un caso de determinado tipo de vehículo computacional (o viceversa) entonces parece que las leyes intencionales no pueden ser implementadas por mecanismos computacionales. Ahora bien, Fodor considera que esto no es un problema si adoptamos una noción más permisiva de implementación. Puede bastar que las propiedades informativas y las computacionales, los contenidos informativos y las formas sintácticas, se correlacionen confiablemente (aunque contingentemente). Luego de plantear este problema y la mencionada solución, Fodor dedica el resto del ensayo a mostrar que efectivamente es probable que haya esta correlación confiable entre propiedades computacionales e informativas.

Si bien en trabajos posteriores Fodor no vuelve a tratar tan específica y exhaustivamente el problema mencionado de la implementación computacional de las leyes intencionales, como veremos más adelante en esta sección, en sus trabajos más recientes aún sostiene las tesis relevantes para la idea que será el foco de este trabajo: que la relación entre contenido informativo y su vehículo computacional es contingente. En lo que sigue centraré el desarrollo de la posición fodoriana en las tesis que, como dije, son relevantes para el objetivo de este capítulo: El enfoque intencional de la cognición, la teoría informativa de la intencionalidad y la caracterización de la relación entre estados informativos y vehículos computacionales.

Fodor sostiene que las explicaciones psicológicas apelan a leyes intencionales. Como dije, no me comprometo en este trabajo con ninguna caracterización de la explicación psicológica, pero sí con la tesis metafísica de que los procesos mentales o cognitivos que causan nuestra conducta involucran muchos estados (como creencias y deseos) que son individuados en virtud de su contenido. Fodor (1994) no da ninguna razón explícita a favor de esta tesis. Sólo afirma que no hay ninguna buena razón en su contra (p. 3). Por mi parte, considero que Fodor tiene buenas razones *para pensar que no hay buenas razones en contra de esta tesis*. Considero que una buena razón para dudar de que existan estados intencionales es que no dispongamos de una caracterización naturalista, en términos físicos, del contenido intencional. Pero Fodor (1994, 1998, 2008) considera que hay una caracterización de este tipo. Esto nos lleva a la siguiente tesis.

Fodor (1994) considera que el contenido intencional está constituido por información. Más explícitamente, Fodor afirma que el contenido de una representación está constituido por la relación que tiene con propiedades u objetos externos al organismo y no por las relaciones que tiene con otras representaciones. Por ejemplo, los pensamientos sobre perros son sobre perros porque son el tipo de

pensamientos que pueden ser causados por perros. Fodor (1998) da una caracterización más clara y específica del enfoque informacional. Según Fodor (1998) las teorías informacionales sostienen que la relación entre los pensamientos y el mundo que determina el contenido de los primeros es una relación nomológica. Dicho de modo general, una representación R tiene a una propiedad P como contenido sólo si hay una ley que determina que las instanciaciones de P causan instanciaciones de R (en ciertas condiciones C a especificar) (Fodor 1998, p. 12). Como vimos en el capítulo 6, la determinación del contenido informacional requiere además un elemento etiológico o teleológico. Pero, como veremos, esta diferencia no es relevante para el problema que nos concierne.

Hay dos razones centrales que ofrece Fodor para aceptar una teoría informacional. La primera es que satisface la exigencia de naturalizar el contenido intencional. Fodor considera que cualquier psicología intencional científicamente aceptable debe dar condiciones naturalistas (es decir, especificables en términos físicos) suficientes para que se instancie un estado intencional. Fodor sostiene que la naturalizabilidad es un requisito de cualquier ciencia no básica y se funda en la creencia de que todo aquello de lo que hablan las ciencias es físico. La propuesta informacional, al reducir el contenido a relaciones nomológicas entre estados cerebrales y propiedades del entorno, parece satisfacer este requisito. Considero que este punto es importante, ya que nos muestra que esta segunda tesis fodoriana es una razón para aceptar la primera. Podemos afirmar que los procesos cognitivos consisten en la manipulación de estados intencionales porque tenemos una caracterización naturalista del contenido intencional.

Hay una segunda razón de Fodor para sostener una teoría informacional, que mencionaremos pero no profundizaremos, ya que nos hemos ocupado extensamente de ella en el primer parte. Una ventaja crucial de las propuestas informacionales, para Fodor, es el hecho de que son atomistas. Si el contenido de una representación R está determinado sólo por la relación que tiene con una propiedad distal y no por las relaciones que tiene con otras representaciones, entonces un sujeto puede poseer R aún cuando no posea ninguna otra representación. Ser atomista es una ventaja porque, como vimos, la teoría alternativa, el funcionalismo, adolece del holismo semántico. Hemos visto en los capítulos 2 y 3 que podemos incorporar roles funcionales sin caer en los problemas que conlleva el holismo. Sin embargo, vimos también que es crucial para evitar estos problemas la posibilidad de un contenido que, como el informacional, no dependa de otras representaciones. Como vimos en el capítulo 2, este es el único modo de que el proceso de ramificación no haga a los roles funcionales demasiado amplios. De este modo, el atomismo informacional sigue siendo, aún desde mi perspectiva, una ventaja crucial de

las propuestas informacionales.

Además, podemos considerar que la propuesta informacional es naturalista en otros dos sentidos, diferentes al considerado por Fodor. En primer lugar respeta el uso de “cognición” por parte de las ciencias encargadas del estudio de los procesos cognitivos. Como dijimos en el capítulo pasado, Piccinini y Scarantino (2010) afirman que los científicos cognitivos asumen usualmente que la cognición involucra el procesamiento de información e identifican a las representaciones con señales que portan (o tienen la función de portar) información (p.24). En segundo lugar, independientemente de su uso por científicos cognitivos, la noción misma de información es una adaptación de una noción desarrollada matemáticamente de modo independiente (a partir de la teoría de Shannon). Considero que estos son dos elementos que le dan a la propuesta informacional una credibilidad adicional desde la perspectiva naturalista.

En trabajos más recientes Fodor ya no tiene la misma actitud hacia las teorías informacionales del contenido. Por ejemplo, Fodor (2008) no se compromete específicamente con una teoría informacional sino que considera, de modo más indeterminado, que alguna teoría causal debe ser verdadera. Las teorías causales del contenido mental son una familia de teorías muy diversas (entre las que se encuentra la teoría informacional) cuyo denominador común es la idea de que el contenido de una representación mental es determinado por algún tipo de relación causal que guarda con un objeto o propiedad distal (Adams y Aizawa 2010). La expresión “teoría causal” es equívoca ya que, por ejemplo, la semántica de rol funcional en su variante naturalista también sostiene que las representaciones adquieren su contenido por medio de relaciones causales. La diferencia es que en el caso del funcionalismo se trata centralmente de relaciones causales entre representaciones. Sin embargo, esta es la terminología estándar (Adams y Aizawa 2010) y Fodor (2008) también la emplea. El punto que quisiera señalar aquí es que, como veremos más adelante, este compromiso más indeterminado de Fodor (2008) con alguna teoría causal del significado es suficiente para plantear la tesis sobre la relación entre la cognición y su base computacional que será relevante en este trabajo. Como veremos un poco más adelante, lo importante para la tesis fodoriana es que el contenido es determinado por alguna propiedad *extrínseca* de la representación, por una relación causal (no importa específicamente cuál) que guarda con un estímulo distal.

La otra tesis fodoriana que quiero considerar es la idea de que las leyes psicológicas son típicamente implementadas por procesos computacionales. Fodor sostiene que tiene que haber un

mecanismo implementador para cualquier ley de una ciencia no básica. Un mecanismo que implementa una ley, según Fodor, es aquel en virtud de cuya operación la satisfacción del antecedente de la ley produce la satisfacción de su consecuente (1994, p. 8). Típicamente, aunque no necesariamente, los mecanismos que implementan una ley son especificados en términos de una ciencia de nivel más bajo.

Los procesos computacionales, según Fodor (1994), son definidos sobre objetos sintácticamente estructurados. Son mapeos de objetos sintácticamente estructurados a objetos sintácticamente estructurados. Más adelante haremos algunas precisiones sobre la noción de estructura sintáctica. Por ahora basta señalar que Fodor (1994) considera que hay un argumento sumamente convincente para considerar que las leyes psicológicas son implementados por procesos computacionales así descriptos. Fodor considera que es característico de los procesos mentales gobernados por estas leyes conservar propiedades semánticas como la verdad. Si alguien comienza con un pensamiento verdadero y lleva a cabo algunos razonamientos, muy frecuentemente llegará a formar otros pensamientos verdaderos. Fodor (1994) considera que este es uno de los hechos más importantes que sabemos sobre las mentes y que por lo tanto cualquier teoría psicológica tiene que dar cuenta de él. Ahora bien, siguiendo a Turing, Fodor considera que si uno tiene un mecanismo cuyas operaciones consistan en transformaciones de símbolos y cuyos cambios de estado estén determinados por las propiedades sintácticas de esos símbolos, entonces es posible organizar un mecanismo de este tipo tal que, en una gran cantidad de casos, el mecanismo transforma confiablemente un input de símbolos verdaderos en un output de símbolos también verdaderos. Fodor considera que esta es la única alternativa relevante para explicar la preservación de verdad.

Como vimos, Fodor (1994) considera que la relación entre las leyes intencionales y la computación es una de implementación. Como veremos, Fodor distingue entre diferentes modos de implementación: por constitución y por realización. A su vez, presenta una noción estándar de implementación y le opone una noción propia. Fodor muestra que la noción estándar de implementación requiere que haya una relación conceptual o metafísicamente necesaria entre las propiedades implementadoras y las implementadas. Sin embargo, argumenta la relación entre los vehículos informacionales y computacionales es contingente. A partir de esto, en lugar de rechazar la tesis de que la cognición es implementada por la computación, Fodor propone un enfoque alternativo de la implementación. En este capítulo no me comprometo con ningún tipo de relación metafísica específica entre cognición y computación. En particular no me comprometo con ninguno de los dos enfoques sobre la implementación. Para ello tenemos que esperar al capítulo 9. Por ahora, me interesa

determinar si es cierta la tesis de que *la instanciación de vehículos informacionales no implica la instanciación de vehículos computacionales* y considerar ciertas consecuencias de esta tesis. Para presentar el argumento a favor de esta relación contingente entre los dos tipos de vehículos desarrollaré brevemente la discusión sobre la implementación en la que aparece.

Fodor (1994) considera que una teoría estándar de la implementación explica cómo los *F*s causan *G*s determinando el mecanismo que la instanciación de un *F* es suficiente para poner en funcionamiento y cuyo funcionamiento causa un estado de cosas que es suficiente para producir una instancia de *G*. Por ejemplo, según Fodor, una teoría de la implementación explica cómo los *F*s causan *G*s, mostrando que la instanciación de un *F* es suficiente para instanciar un M_F , que M_F es suficiente para instanciar M_G y, a su vez, M_G es suficiente para instanciar *G*. Por medio de las relaciones entre *F*, M_F , M_G y *G* podemos mostrar cómo los *F*s causan *G*s. Fodor considera, sin embargo, que este enfoque sobre la implementación, tal como fue descrito, no permite defender la idea de que las leyes intencionales son implementadas por procesos computacionales.

Fodor considera que hay dos relaciones que pueden mantener las leyes y sus mecanismos implementadores: reducción y realización múltiple. Ahora bien, parece que las leyes psicológicas, tal como las caracterizó Fodor no guardan ninguna de estas relaciones con los mecanismos computacionales. La reducción se da siempre que una propiedad de nivel N es identificable con una propiedad de nivel N-1 (por ejemplo, comúnmente se cree que ser agua es igual a ser H₂O). La realización múltiple se da siempre que hay una disyunción de propiedades N-1 tal que: (I) la instanciación de cualquiera de los disyuntos es suficiente para que se instancie la propiedad de nivel N y (II) la instanciación de la propiedad de nivel N es suficiente para la instanciación de la disyunción (aunque no para la de alguno de los disyuntos en particular). Tanto las nociones de reducción como las de realización múltiple, tal como las caracteriza Fodor, determinan que la instanciación de una propiedad de nivel N es *suficiente* para la instanciación de alguna propiedad de nivel N-1 y la instanciación de una propiedad de nivel N-1 es *suficiente* para la instanciación de una propiedad de nivel N.

Sin embargo, Fodor cree que no hay estados computacionales suficientes para la instanciación de un estado intencional (informacionalmente caracterizado) y viceversa, no parece haber estados intencionales suficientes para la instanciación de un estado computacional determinado (o una disyunción de estados computacionales determinados). Fodor considera que esto se sigue de que los

estados computacionales son caracterizados por su estructura sintáctica y las formas sintácticas son propiedades *intrínsecas*. Por el contrario, los estados informacionales son caracterizados por sus relaciones causales con objetos distales, es decir, por sus relaciones *extrínsecas*. Fodor cree que si bien puede haber correlaciones empíricas entre propiedades intrínsecas y extrínsecas, tener una propiedad extrínseca no es una condición *metafísicamente* suficiente para tener ciertas propiedades intrínsecas, ni viceversa. Para cualquier par de propiedades, una intrínseca *I*, y otra extrínseca *E*, existe un mundo posible donde un objeto tiene *E* pero no *I* y otro donde tiene *I* pero no *E*. Por lo tanto, *la relación entre las representaciones que figuran en las leyes intencionales y los vehículos computacionales es metafísicamente contingente*. En el capítulo que sigue pondré en duda esta última premisa, pero por el momento lo importante es notar que si la aceptamos, y si aceptamos que la implementación requiere una relación necesaria entre estos dos elementos, entonces las leyes intencionales no son implementadas por la computación. Como vimos, este punto se sigue incluso si, como Fodor (2008), no sostenemos específicamente una teoría informacional sino algún otro tipo de teoría causal. Pues lo relevante es que el contenido representacional de una representación *R* esté determinado por alguna relación extrínseca de *R* con un estímulo distal.

Brevemente, la solución fodoriana a este problema es la siguiente. Aunque la relación entre los estados representacionales y computacionales es metafísicamente contingente, no necesitamos una relación metafísicamente necesaria entre ellos para que la segunda implemente a la primera. Basta con que estos dos elementos estén correlacionados confiablemente, en el rango de los mundos cercanos. Basta con que la base computacional se instancie en los mundos cercanos en los que se instancia su contenido informacional y el contenido informacional se instancie en los mundos cercanos en los que se instancia su base computacional. En el resto del trabajo, Fodor (1994) se dedica a mostrar que efectivamente se da esta correlación confiable entre estados informacionales y computacionales.

3. Cognición y sintaxis

Habiendo distinguido las tesis y la argumentación fodoriana relevante para la idea de que la cognición y la computación se relacionan contingentemente, en esta sección quiero distinguir diferentes modos de entender dichas tesis y evaluar los argumentos considerados a la luz de estas diferentes lecturas. En primer lugar, hay al menos dos modos de entender una tesis sobre la relación entre

computación y cognición. En segundo lugar, hay diferentes modos en que podemos entender la noción fodoriana de computación. Empecemos por la primera distinción.

Hay dos tesis diferentes, aunque vinculadas, que podemos sostener sobre la relación entre estados informacionales y formas sintácticas. Podemos sostener que en todos los mundos en los que se instancia un estado informacional se instancia una forma sintáctica cualquiera, y viceversa. Es decir, no puede instanciarse un estado informacional sin que se instancie *alguna* forma sintáctica (y viceversa). Pero también podemos sostener algo más específico sobre la relación entre lo informacional y lo sintáctico. Podemos sostener que, para todo *tipo específico* de estado informacional $I_1, I_2, I_3, \dots, I_n$, hay un *tipo específico* de forma sintáctica S_1, S_2, S_3, \dots ó S_n , o una disyunción de tipos específicos de formas sintácticas $S_{1.1}, S_{1.2}, S_{1.3}, \dots$ ó $S_{2.1}, S_{2.2}, S_{2.3}, \dots$ ó $S_{n.1}, S_{n.2}, S_{n.3}, \dots$, tal que en todos los mundos en los que se instancia I_1 se instancia S_1 (o la disyunción de $S_{1.1}, S_{1.2}, S_{1.3}, \dots$), y viceversa, en todos los mundos en los que se instancia I_2 , se instancia S_2 (o la disyunción de $S_{2.1}, S_{2.2}, S_{2.3}, \dots$) y viceversa, y en todos los mundos donde se instancia I_n , se instancia S_n (o la disyunción $S_{n.1}, S_{n.2}, S_{n.3}$), y viceversa. La primera tesis afirma una relación necesaria *general* entre estados informacionales y formas sintácticas. La segunda tesis afirma una relación necesaria entre *tipos específicos* informacionales y computacionales. Por este motivo, las llamare “*correlación general*” y “*correlación específica*”, respectivamente. Considero que *correlación general* es de particular interés porque, si fuese falsa, sería cierto que la instanciación de un estado informacional no sólo no implicaría la instanciación de una forma sintáctica de este o aquél tipo, sino que no implicaría la instanciación de ninguna forma sintáctica en absoluto y por lo tanto la instanciación de un proceso cognitivo no implicaría la de uno computacional.

Ambas tesis están vinculadas. *Correlación específica* implica *correlación general*. Si no puede instanciarse un estado informacional de determinado tipo sin que se instancie una forma sintáctica de otro tipo determinado (y viceversa), entonces, *ipso facto*, no puede instanciarse un estado informacional cualquiera sin que se instancie también una forma sintáctica cualquiera (y viceversa). Sin embargo, las tesis no son equivalentes. *Correlación general* no implica *correlación específica*. Puede darse el caso de que, necesariamente, no se instancia un estado informacional sin que se instancie una forma sintáctica cualquiera (y viceversa), pero que no en todos los mundos en los que se instancia un estado informacional de *determinado tipo* I_1 se instancia una forma sintáctica *de un mismo tipo determinado* S_1 (o una forma sintáctica que caiga bajo alguno de los disyuntos $S_{1.1}, S_{1.2}, S_{1.3}, \dots$). Es decir aún si *correlación general* es verdadera, aún si no se puede instanciar I_1 sin que se instancie alguna forma sintáctica, esta forma puede no ser siempre del mismo tipo (o perteneciente a alguno de una disyunción

de tipos).

Es claro que Fodor argumenta explícitamente en contra de *correlación específica*. Fodor quiere explicar cómo un estado informacional específico I_1 causa otro estado informacional específico I_2 mostrando que I_1 y I_2 son implementados por formas sintácticas *específicas* S_1 y S_2 . El problema de Fodor es que la noción estándar de implementación requiere que para que I_1 y I_2 sean implementados por S_1 y S_2 la instanciación de I_1 debe implicar necesariamente la instanciación de S_1 y la instanciación de S_2 debe implicar necesariamente la de I_2 . Pero, como argumenta Fodor, ningún estado informacional *específico* implica necesariamente la instanciación de una forma sintáctica *específica*, ni viceversa. Esta es justamente la negación de *correlación específica*.

Ahora bien, dado que *correlación general* no implica *correlación específica*, la negación fodoriana de *relación específica* no implica la negación de *relación general*. Sin embargo, podría pensarse que las razones de Fodor para rechazar *correlación específica* implican también la negación de *correlación general*. Pues si es cierto que, como afirma Fodor, la posesión de una propiedad extrínseca no implica necesariamente la posesión de una propiedad intrínseca, ni viceversa, entonces la instanciación de un tipo específico estado informacional no sólo no implica la instanciación de un tipo *específico* de forma sintáctica sino tampoco *la instanciación de forma sintáctica alguna*. Pues tanto tener una forma sintáctica de un tipo específico como tener una forma sintáctica *en general* son propiedades intrínsecas. Tener una forma sintáctica *en general* no implica ni es implicado por la posesión de ninguna propiedad informacional.

Sin embargo, creo que si consideramos detalladamente la noción fodoriana de sintaxis, notaremos este argumento en contra de *correlación general* no funciona. Vimos que según Fodor (1994) y (1998) la computación es el procesamiento de símbolos en función de sus propiedades sintácticas. Cuando el símbolo es complejo, su estructura sintáctica está determinada por la forma sintáctica de cada uno de sus componentes. Las formas sintácticas son las propiedades a las que responden, según Fodor, los procesos computacionales en general, y los procesos cognitivos en particular ¿Pero qué tipos de propiedades constituyen formas sintácticas? Fodor (2008) afirma que *cualquier propiedad física* puede constituir una forma sintáctica. Fodor afirma que la forma sintáctica puede ser cualquier propiedad física de las representaciones a las que el procesamiento cognitivo responda (p. 79).

Ahora bien, una vez que caracterizamos a la computación como manipulación de formas

sin-tácticas y afirmamos que cualquier propiedad física puede ser una forma sin-táctica se sigue *correlación general*. Pues dado que los estados informacionales son propiedades extrínsecas de *objetos físicos*, objetos que por definición tienen propiedades físicas intrínsecas, la instanciación de un estado informacional implica la instanciación de *alguna* propiedad física intrínseca. Sin embargo, claramente, *correlación general* se sigue sólo porque la noción de computación fodoriana es trivial. No permite distinguir un vehículo computacional de cualquier otro tipo de estado físico.

Es claro que si queremos ofrecer una noción no trivial de forma sin-táctica tampoco podemos identificar a las formas con propiedades neuronales. Pues los vehículos computacionales pueden realizarse en bases no neuronales. Si existe algo como la computación digital, por ejemplo, entonces según Piccinini y Bahar (2013), no es algo que se instancie en un cerebro. Sin embargo, esta posibilidad misma de realización en diferentes bases físicas puede darnos la característica más específica que necesitamos. Hemos visto que, a pesar de que Fodor (2008) dice explícitamente que las formas son propiedades físicas cualesquiera, dice algo más sobre ellas en el marco de su tratamiento del problema de la publicidad. Fodor sostiene que no sólo los estados informacionales, *sino también las propiedades computacionales, como la forma sin-táctica*, pueden realizarse en diferentes bases neuronales (p. 90). Como afirma Fodor, es más o menos conocimiento común que existe la posibilidad, al menos en principio, que los estados computacionales sean realizados múltiplemente en diferentes cerebros. Casos de la misma representación mental pueden ser realizados por propiedades sub-computacionales diferentes que son del mismo tipo bajo descripción computacional.

De este modo, las propiedades computacionales no son propiedades físicas cualesquiera, son propiedades físicas, intrínsecas y múltiplemente realizables en estados de nivel inferior. Si bien Fodor no incluye en su caracterización de la forma sin-táctica la propiedad de la realizabilidad múltiple, sino que sólo la menciona en el marco de un argumento, podemos tomarla para ofrecer una noción más robusta que la que propone. Esta noción no es tan trivial como la anterior. No cualquier estado físico es múltiplemente realizable. Sin embargo, esta noción más robusta parece poner nuevamente en funcionamiento el argumento que ofrecí antes en contra *correlación general*. *Correlación general* es falsa porque como la instanciación de un estado informacional no implica la instanciación de propiedad específica intrínseca alguna, no implica, en particular, la instanciación de una propiedad intrínseca múltiplemente realizable. En la sección siguiente evaluaré nuevamente estos argumentos en torno a las dos tesis sobre la *correlación general y específica* entre estados informacionales y computacionales, a la luz de los desarrollos más recientes sobre la noción de computación.

4. Cognición y computación genérica

Piccinini y Bahar (2013) distinguen la noción de computación de otras 4 nociones: Las de (1) sistema de input-output, (2) sistema de input-output funcionalmente organizado, (3) sistema de feedback control y (4) sistema de procesamiento de información. Piccinini y Bahar consideran que las nociones (1) y (2) son demasiado generales como para ser identificadas con la de computación. Un sistema computacional no es meramente un sistema de input-output. Un sistema de input-output es un sistema que genera respuestas a partir del procesamiento de cierto estímulo. Esta noción no determina nada sobre la naturaleza del input, el output o *los estados internos* encargados del procesamiento. Como señalan Piccinini y Bahar, si la noción de input y de output son suficientemente liberales, entonces cualquier sistema físico pertenece a esta clase de sistemas (p. 455). Por lo tanto, no podemos caracterizar a los sistemas computacionales como sistemas de input-output.

La noción de un sistema de input-output funcionalmente organizado es algo más específica que la anterior. Se trata de la idea de un sistema cuya *función* es generar cierto output a partir del procesamiento de cierto input. Pero, como señalan Piccinini y Bahar, esta noción sigue siendo demasiado general como para ser identificada con la de computación. No todos los sistemas funcionalmente organizados son sistemas computacionales. Por ejemplo, los motores, refrigeradores y estómagos son sistemas funcionalmente organizados que intuitivamente no computan (p. 445).

La noción de forma sintáctica que presenta explícitamente Fodor (2008) la idea de que la forma sintáctica es cualquier propiedad física a la que responde un proceso cognitivo, da como resultado una noción de computación tan general como la noción de procesamiento input-output (o, como mucho, con la de procesamiento input-output funcionalmente organizado). La noción de vehículo computacional de Fodor, la noción de forma sintáctica, no nos permite distinguir entre un proceso computacional y cualquier otro tipo proceso físico (o cualquier otro proceso físico funcionalmente organizado). Si la computación es simplemente un proceso en el que las propiedades causalmente relevantes pueden ser propiedades físicas cualquiera, entonces parece que cualquier proceso físico (o tal vez, cualquier proceso físico funcionalmente organizado) puede contar como computacional.

Podría argumentarse que la caracterización fodoriana de la computación en términos de manipulación de *símbolos*, implica que los estados que participan en el procesamiento computacional no tienen sólo una sintaxis sino también una semántica y claramente, no cualquier proceso físico

(funcionalmente organizado o no) tendrá una semántica asociada a su sintaxis. Sin embargo, si aceptamos la caracterización fodoriana del contenido en términos de información, entonces, los vehículos de la computación no necesitan tener contenido. Como señalan Piccinini y Bahar (2013) los vehículos del procesamiento computación no necesitan portar información. Un sistema computacional puede manipular secuencias de dígitos que no portan información sobre algo externo al sistema, según alguna regla arbitraria (p. 457). Por lo tanto, los vehículos computacionales no pueden ser caracterizados en función de un contenido semántico.

Ahora bien, vimos que si bien la noción fodoriana es demasiado general, menciona en el marco de su argumentación en torno al requisito de publicidad otra característica de los vehículos computacionales. Se trata de estados que pueden realizarse en diferentes estructuras neuronales. Si unimos la idea de Fodor (1994) de que las propiedades computacionales son intrínsecas y la de Fodor (2008) de que son propiedades múltiplemente realizables, obtenemos una caracterización de los vehículos computacionales mucho más robusta que la de Fodor (2008), que la idea de que son cualquier propiedad física. La computación opera sobre propiedades intrínsecas que pueden realizarse múltiplemente.

Esta noción es más robusta porque distingue a las propiedades computacionales tanto de propiedades extrínsecas como de propiedades intrínsecas no realizables múltiplemente. Como vimos, esto parece bastar para argumentar en contra de *correlación general*. Si las propiedades informacionales, en tanto extrínsecas, no implican la instanciación de ningún tipo de propiedad intrínseca en particular, no implican tampoco la instanciación de una propiedad intrínseca múltiplemente realizable. Un estado informacional podría ser realizado en un objeto cuyas propiedades intrínsecas no sean múltiplemente realizables.

Es interesante notar que la noción propuesta por Piccinini y Bahar (2013) y presentada en el capítulo pasado es muy semejante a la noción que he presentado aquí, recogiendo elementos de diferentes trabajos de Fodor. Retomemos con un poco de detalle la noción de Piccinini y Bahar. La *computación genérica* es el procesamiento de vehículos de acuerdo con reglas que son sensibles sólo a ciertas propiedades de ellos. Las reglas (los mapas de input-output) que definen la computación genérica son sensibles sólo a dimensiones específicas de variación de los vehículos, no son sensibles a propiedades físicas más específicas o concretas de ellos (p. 458). Piccinini y Bahar llaman a estos vehículos que definen la computación genérica “independientes del medio” ya que al no ser muy

concretas sus propiedades funcionalmente relevantes, pueden ser realizados por medios muy diferentes (mecánico, electromecánico, electrónico, magnético, etc.).

Como vimos en el capítulo pasado, la noción de computación genérica captura los sentidos de 'computación' relevantes. Según Piccinini y Bahar, los dos tipos de computación más generales, la computación digital y la analógica, comparten la característica de la independencia del medio. Se trata, entonces, de una noción que tiene dos ventajas. Por un lado, no es trivial, puesto que puede haber muchos sistemas físicos que, al ser sensibles a propiedades físicas muy específicas de los estados que manipulan, no tendrán vehículos independientes del medio. Por otro lado, se trata de una caracterización que está más estrechamente ligada al fenómeno considerado, puesto surge de la consideración de diferentes tipos de sistemas computacionales.

Es claro que esta noción está vinculada a la que construí a partir de Fodor. Vimos que para Fodor la computación opera sobre propiedades intrínsecas múltiplemente realizables. Si bien Piccinini y Bahar no se definen sobre el carácter intrínseco de las propiedades computacionales, sí afirman que son múltiplemente realizables, que pueden realizarse en diferentes medios. Pero afirman algo más que ello, definen a la computación por *el tipo peculiar de múltiple realizabilidad de los vehículos computacionales*, la independencia del medio. Además, vinculan esta propiedad con los dos tipos más generales de computación. La noción surge del rasgo común de la computación digital y la analógica. Estas dos razones nos muestran que si bien la noción fodoriana podría estar encaminada, la de computación genérica es más precisa y adecuada. Ahora bien, considero que hay una tercer razón que muestra que la noción de computación genérica no es meramente más precisa que la de Fodor, sino además incompatible con ella. Vimos en el capítulo pasado que la noción de computación genérica implica que el procesamiento neuronal no implemente o realiza a la computación sino que es simplemente un tipo específico de computación. Vimos que esto era así porque la noción de computación genérica se aplica al procesamiento neuronal en exactamente el mismo sentido en que se aplica a las nociones de computación digital y analógica. Si esto es así, entonces se sigue que los vehículos computacionales del cerebro no pueden caracterizarse en base a su realizabilidad múltiple en diferentes bases neuronales. Por estos motivos considero que no podemos conservar la noción fodoriana y debemos adoptar la de Piccinini y Bahar.

Como vimos en el capítulo pasado, basados en la noción de computación genérica, Piccinini y Bahar (2013), ofrecen dos argumentos a favor de lo que antes caracteriza como la primer tesis

computacionalista, la idea de que el procesamiento neuronal es computacional. El primer argumento consistió en mostrar que los vehículos del procesamiento neuronal son independientes del medio. Como los aspectos funcionalmente relevantes de los vehículos del procesamiento neuronal (cambios en la tasa y la duración de los potenciales de acción) pueden ser implementados por medios físicos muy diferentes (como tejido neuronal o circuitos de silicona) (Craver 2010), parece que estos vehículos son independientes del medio. Por lo tanto el procesamiento neuronal es computacional en el sentido genérico.

El segundo argumento que Piccinini y Bahar (2013) ofrecen resulta de particular interés para este capítulo ya que se basa en una relación conceptual que guardan, según los autores, las nociones de información y computación genérica. Hemos visto que para Fodor no hay una relación conceptual o metafísica entre los estados informacionales y los vehículos computacionales. Su relación puede ser confiable, pero es metafísicamente contingente. Piccinini y Bahar consideran que de la noción de computación que ellos proponen se sigue una caracterización diferente de la relación entre computación e información. Como vimos, afirman que los vehículos de un sistema computacional pueden no poseer estados informacionales, pueden no portar información. Un computador digital, por ejemplo, puede manipular secuencias de dígitos sin que estas secuencias porten información sobre nada externo al sistema. Sin embargo, consideran que todo sistema que manipule información será un sistema computacional en el sentido genérico. Piccinini y Bahar argumentan que procesar información es procesar vehículos de acuerdo con la información que portan, independientemente de sus propiedades físicas. La información que porta una señal es definida independientemente de las propiedades físicas de dicha señal (como vimos, es algún tipo de correlación entre la señal y un estímulo externo al sistema). De este modo, las reglas del procesamiento de información son definidas de un modo independiente del medio. Por lo tanto, los vehículos del procesamiento de información son computacionales en el sentido genérico. Piccinini y Bahar concluyen que como los vehículos neuronales portan información que es procesada por el sistema nervioso se sigue que el procesamiento neuronal es computacional. Dado que acepto, en base a consideraciones fodorianas y propias, una caracterización de la noción de cognición en términos de procesamiento de información, parece que, en contra de lo que argumentó Fodor, si aceptamos la caracterización de Piccinini y Bahar de la noción de computación, deberíamos concluir a partir de este argumento que cognición implica computación.

Antes de considerar este argumento, es importante señalar el tipo de relación entre información y computación que Piccinini y Bahar consideran. Hemos visto en la sección anterior que hay dos tesis

diferentes que podemos sostener sobre la relación entre lo informacional y lo computacional. Podemos afirmar que hay una correlación necesaria de *tipos específicos* informacionales y computacionales o que hay una correlación general necesaria entre estados informacionales y computacionales. Llamé a estas tesis “*correlación específica*” y “*correlación general*” respectivamente. Vimos en la sección anterior que Fodor argumenta explícitamente en contra de *correlación específica* y que si tomamos la noción de computación de Fodor, se sigue *correlación general*, pero que, al ser la noción fodoriana de computación trivial, no resulta de interés (equivale a la tesis de que los estados informacionales son necesariamente estados físicos). Sin embargo, vimos que si construimos una noción de computación a partir de lo que en diferentes momentos Fodor dice sobre las formas sintácticas, podemos ofrecer una noción no trivial, pero de la que no se sigue *correlación general*.

Ahora bien, considero que el argumento de Piccinini y Bahar constituye un argumento a favor de cierta forma de correlación general, (entendida, por supuesto, no como una tesis sobre formas sintácticas sino sobre vehículos de la computación genérica). Así Piccinini y Bahar argumentarían a favor de *correlación general* empleando una noción no trivial de computación. Piccinini y Bahar argumentan que ser un estado informacional en general implica ser un estado computacional en general. Esto no es exactamente *correlación general* ya que esta tesis afirma una implicación en ambas direcciones, y Piccinini y Bahar niegan explícitamente que computación implique información. Sin embargo, afirman una relación general de implicación de información a computación. Ser un estado informacional cualquiera implica ser un vehículo computacional cualquiera.

Vimos que el argumento fodoriano en contra de *correlación general* parte de la premisa de que la instanciación de una propiedad intrínseca no implica la instanciación de una propiedad extrínseca ni viceversa. Dado que las propiedades informacionales y las computacionales son, respectivamente, según Fodor, extrínsecas e intrínsecas, la instanciación de las primeras no implica la instanciación de las segundas, ni viceversa. Piccinini y Bahar, por su parte, aunque admiten que las propiedades informacionales y las computacionales son diferentes (para ellos, por ejemplo, ser un estado computacional no implica ser un estado informacional), aún si difieren en el modo en que Fodor señala, consideran que estos dos tipos de estado comparten una característica (la independencia del medio) y que esta característica es suficiente para que haya una relación de implicación de lo informacional a lo computacional.

Como vimos en el capítulo pasado, considero que el argumento no funciona. Mostré que, en

contra de Piccinini y Bahar, no hay una relación de implicación de información a computación genérica. Los estados informacionales pueden tener vehículos que no sean independientes del medio en el sentido computacional. He mostrado que los vehículos computacionales e informacionales tienen diferentes tipos de realizabilidad múltiple. Los vehículos computacionales se realizan múltiplemente porque tienen relevancia funcional restringida y los vehículos informacionales porque son definidos por propiedades relacionales.

Como vimos, al ser definido por sus poderes causales, las propiedades funcionalmente relevantes para esos poderes de un caso de un vehículo informacional dado pueden ser de cualquier tipo. Pero esto quiere decir que no es necesario que las propiedades funcionalmente relevantes de un caso de vehículo informacional sean restringidas a aspectos más abstractos. Cualquier tipo de vehículo informacional puede ser realizado por un caso cuyas propiedades funcionalmente relevantes sean muy específicas o concretas. Por lo tanto, ser informacional no implica tener relevancia funcional restringida, no implica ser computacional.

Ahora bien, la relación de implicación considerada hasta ahora es, como vimos, la que involucra *correlación general*, una correlación general entre lo informacional y lo computacional ¿Qué sucede con *correlación específica*, si tomamos la noción de computación genérica? Es claro que tampoco información implicará computación. Pues vimos que *correlación específica* implica *correlación general* y que por lo tanto la negación de *correlación general* implica la negación de *correlación específica*. Ahora bien, acabo de argumentar que *correlación general* interpretada según la noción de computación genérica, es falsa. Por lo tanto, también es falsa *correlación específica*. Si, como argumenté, los estados informacionales pueden realizarse en vehículos que no sean vehículos de la computación genérica entonces, *ipso facto*, ser un estado informacional de un tipo específico I_1 no implica ser un vehículo computacional específico de un tipo S_1 (pues no implica ser un vehículo computacional de ningún tipo). Por lo tanto, la instanciación de un tipo específico de estado informacional no implica la instanciación de un tipo específico de estado computacional.

5. La relevancia cognitiva de la sintaxis

Hemos visto que la caracterización que ofrece Fodor de la noción de forma sintáctica es

demasiado general como para caracterizar a la computación. Vimos que hay una noción más robusta que podemos formar a partir de diferentes propiedades que Fodor atribuye ocasionalmente a las formas sintácticas. Sin embargo, la noción de Piccinini y Bahar no sólo es más adecuada (al uso de “computación” por parte de las ciencias de la computación) sino que es incompatible con esta noción más robusta. Ahora bien, considero que aunque la noción *más amplia* de forma sintáctica no nos sirve para caracterizar a la computación, sí puede decirnos algo importante sobre el procesamiento cognitivo.

Vimos que Fodor afirma que la forma sintáctica es cualquier propiedad física a la que responden los procesos cognitivos. Los procesos cognitivos son, como vimos, operaciones sobre representaciones. Más aún, son procesos que son sensibles al contenido de las representaciones. Es decir, son procesos que responden de modo diferente a representaciones que tienen diferente contenido y de modo semejante a representaciones que tienen contenido diferente. Si adoptamos una teoría informacional de las representaciones, esto equivale a decir que los procesos cognitivos son procesos que responden de modo semejante a estructuras que portan (o tienen la función de portar) la misma información y de modo diferente a estructuras que portan (o tienen la función de portar) diferente información. Ahora bien, si reparamos en ciertas cosas que ya sabemos sobre los estados informacionales, notaremos que nos hace falta algo más que estados informacionales para que se de la cognición en este sentido.

Vimos que la instanciación de un tipo informacional no implica la instanciación de ninguna propiedad intrínseca específica. Vimos que tanto un patrón de rayos de luz como un patrón de movimiento molecular pueden portar la información de que la presa está presente. Ahora bien, esto implica que el hecho que un estado porte cierta información *no implica que tendrá ciertos poderes causales específicos*. Por ejemplo, el patrón de rayos de luz y no el patrón de movimientos moleculares serán capaces de activar el procesamiento visual en un humano normal (y el patrón de movimientos pero no el de rayos activará el sistema auditivo). Estados que portan la misma información pueden tener poderes causales diferentes. Dándole un sentido diferente al de Fodor, aunque relacionado, llamaré “forma sintáctica” a la propiedad intrínseca de un estado que porta información y que determina sus poderes causales. Este sentido es diferente al de Fodor porque, aunque Fodor considera que la forma sintáctica determina los poderes causales de lo que la posee (vimos esto en el capítulo 2), como Fodor quiere que la forma sintáctica caracterice a la computación, es claro que debe poder darse en estados que no porten información. Por el contrario, creo que la forma sintáctica puede ser más iluminadora si la consideramos exclusivamente como propiedades de estados con contenido

informativa. Las formas sintácticas son las propiedades físicas intrínsecas de un estado informativo que determinan los poderes causales del mismo.

Decir, como supusimos, que el procesamiento cognitivo requiere que, generalmente, se responda a diferentes piezas de información de modo diferente y de modo semejante a piezas de información del mismo tipo quiere decir que el procesamiento cognitivo requiere que estados que portan información diferente tengan poderes causales diferentes y estados que portan la misma información tengan poderes causales semejantes. Pero si lo que determina los poderes causales de un estado informativo es su forma sintáctica, sus propiedades físicas intrínsecas, esto quiere decir que para que haya cognición debe haber una correlación entre la información que portan los estados sobre los que opera el procesamiento y las propiedades intrínsecas de esos estados. Como estas propiedades determinan sus poderes causales, el procesamiento de información sólo puede tener lugar si estados con diferente información tienen, generalmente, diferentes formas sintácticas y estados con la misma información tienen la misma forma sintáctica. Si bien la instanciación de un tipo informativo no implica la instanciación de una propiedad intrínseca, la instanciación de un *proceso* informativo o cognitivo implica algo sobre las propiedades intrínsecas de los estados informativos. Implica que hay una correlación confiable (aunque contingente) entre tipos de propiedades intrínsecas y tipos de estados informativos. Si bien *correlación específica*, referida a formas sintácticas, es falsa, es cierto el condicional que tiene como antecedente la instanciación de un proceso informativo o cognitivo y como consecuente a *correlación específica* (referida a las formas sintácticas del proceso informativo instanciado).

Obviamente, las formas sintácticas entendidas de este modo no necesitan ser computacionales. Es decir, no es cierto el condicional de que si se instancia un proceso informativo, se da, respecto de ese proceso, una correlación entre tipos informativos y computacionales. Las formas sintácticas de dicho proceso podrían ser propiedades físicas muy específicas que no sean independientes del medio. Lo único que necesitamos es que generen las diferencias y semejanzas de poderes causales necesarias para el procesamiento de información. Así, si bien el procesamiento de información si requiere algo sobre las propiedades intrínsecas de los estados que manipula, no requiere que estos sean computacionales. Al igual que la noción de información, el procesamiento de información tampoco implica computación. Sin embargo, las formas sintácticas nos dicen algo importante sobre la relación entre computación y cognición. Aún si aceptamos a partir del argumento de Piccinini y Bahar (2013) que el procesamiento neuronal es computacional, y aceptamos, a partir de la evidencia citada por ellos

mismos, que el procesamiento neuronal es procesamiento de información y que por lo tanto, aunque computación no implique información ni viceversa, de modo contingente, los estados neuronales (o al menos algunos de ellos) son tanto computacionales como informacionales, *no se sigue que la computación neuronal sea informacional o cognitiva*. Veamos por qué.

Para que la cognición sea computacional es necesario que las formas sintácticas sean computacionales. Es decir, *hace falta que las propiedades intrínsecas que hacen posible que el procesamiento neuronal responda de modo diferente a señales que portan información diferente sean computacionales*. Pero para ello es necesario que haya una *correlación entre diferentes tipos de información y diferentes tipos de vehículos computacionales*. El argumento de Piccinini y Bahar determina meramente que el procesamiento neuronal responde a un tipo general de vehículo, las tasas y duraciones de los potenciales de acción, que son independientes del medio. Además, Piccinini y Bahar citan evidencia respecto de que estados neuronales portan y procesan información sobre estados externos. Para determinar que el procesamiento cognitivo, que el procesamiento de información es computacional hace falta determinar aún algo más. Hay que determinar si es cierta la siguiente tesis:

Computación neuronal cognitiva: *Las diferentes tasas y tiempos de potencial de acción son las propiedades intrínsecas de los estados neuronales que portan diferentes piezas de información responsables de que el procesamiento neuronal responda de modo diferente a esas diferentes piezas de información.*

Si no son estas diferencias entre vehículos neuronales, sino algo más, lo que genera las diferencias causales requeridas para el procesamiento de información, entonces las tasas y tiempos de los potenciales no son las formas sintácticas del procesamiento cognitivo, el procesamiento cognitivo no opera sobre tasas y tiempos de potenciales. Por lo tanto, aunque las tasas y potenciales sean independientes del medio y aunque porten información, no se sigue que la cognición sea computacional.

Por ejemplo, supongamos que dos poblaciones neuronales n y n_1 portan información sobre diferentes estímulos externos, respectivamente, e y e_1 . Es decir, n y n_1 están en una relación nomológica con e y e_1 , respectivamente. Parte de la respuesta que determina estas relaciones nomológicas es la tasa

y tiempo de disparo d y d_1 de n y n_1 en presencia de e y e_1 . Tanto n como n_1 adquieren cierta tasa y/o tiempo de disparo específica d y d_1 en presencia de e y e_1 . Ahora bien, supongamos que d y d_1 tienen el mismo valor, son la misma tasa y/o tiempo de disparo. Supongamos además que, sin embargo, los potenciales de n y n_1 producen diferentes respuestas por parte del sistema nervioso. Si esto es así, entonces no son las tasas y tiempos de los potenciales los vehículos del procesamiento de información neuronal, sino alguna otra propiedad de los potenciales que sea responsable de su diferencia causal.

6. Conclusión

En este capítulo he mostrado principalmente dos cosas: Que no hay una relación conceptual entre las nociones de cognición y computación y que para que la cognición sea computacional debe haber una correlación específica entre tipos computacionales e informacionales en un proceso cognitivo dado. La primera tesis fue defendida por Fodor, pero sobre la base de una noción inadecuada de computación y basado en una premisa falsa sobre la relación entre las propiedades intrínsecas y extrínsecas. La misma tesis fue rechazada por Piccinini y Bahar sobre la base de una noción adecuada de computación, pero una caracterización inadecuada de la relación entre computación e información. Propuse una defensa de dicha tesis basado en la noción de Piccinini y Bahar y mostré las ventajas de mi argumentación respecto de la de Piccinini y Bahar (2013).

Respecto de la segunda tesis, mostré que si bien la noción de forma sintáctica no nos sirve para caracterizar a la computación, nos muestra un aspecto crucial de la cognición entendida como procesamiento informacional. El procesamiento informacional requiere que haya una correlación entre tipos informacionales y los tipos de propiedades intrínsecas de los estados informacionales, las propiedades que determinan sus poderes causales. Esta idea implica que hay una restricción importante para que un proceso cognitivo sea computacional. Los diferentes tipos de información que portan los estados neuronales deben estar correlacionados con los diferentes tipos de vehículo de la computación neuronal.

A pesar de este hiato conceptual entre la computación neuronal y la cognición, considero que hay un vínculo metafísico entre el procesamiento neuronal y la cognición. Creo que hay aspectos no computacionales del procesamiento del sistema nervioso que son cruciales para caracterizarlo y que a

la vez trazan un vínculo estrecho entre las nociones de procesamiento neuronal y procesamiento de información. En el próximo capítulo me detendré a considerar la relación entre la función misma del procesamiento neuronal, el control por retroalimentación negativa, y el procesamiento de información.

Capítulo 9:

Procesamiento neuronal y cognición

1. Control por retroalimentación negativa

Vimos que Piccinini y Bahar (2013) ofrecen una muy buena caracterización de la noción de computación. A su vez, consideran que hay una relación conceptual entre esta noción y otras dos nociones vinculadas a la tesis computacionalista. Consideran que la noción de computación genérica es implicada tanto por la noción de procesamiento neuronal como por la noción de información. Podemos preguntarnos qué determinan estas relaciones conceptuales si, siguiendo a Fodor, adoptamos una concepción informacional de la cognición. Es claro que estas relaciones no establecen una relación conceptual entre el procesamiento neuronal y la cognición. Aunque información y procesamiento neuronal impliquen computación, pueden no implicarse entre sí. Por ejemplo, el procesamiento neuronal y el procesamiento de información podrían ser dos tipos diferentes de computación. Sin embargo, se sigue de la caracterización de Piccinini y Bahar que los procesos neuronales y los cognitivos son al menos procesos de un mismo tipo general, ambos son procesos computacionales.

En el capítulo 7 argumenté que aunque la noción de procesamiento neuronal implica la de computación, la de información no lo hace. En el capítulo 8 mostré como funciona esta última idea a favor de la tesis de que la cognición no implica computación (ni viceversa). Esto hace el hiato entre procesamiento neuronal y cognición aún más grande de lo que era según la elucidación de Piccinini y Bahar. Pues ahora ni siquiera podemos decir que la cognición es un tipo de proceso computacional. No sólo no hay una relación de implicación entre las nociones de procesamiento neuronal y cognitivo sino que ni siquiera se trataría de procesos de un mismo tipo general.

Ahora bien, en esta sección mostraré que a pesar de este hecho, el procesamiento neuronal y la

cognición están aún más estrechamente relacionados que en la caracterización de Piccinini y Bahar. La caracterización de la noción de procesamiento neuronal de estos autores está centrada en el tipo de vehículo que este tipo de proceso manipula (tasas y tiempos de potenciales de acción neuronal). Sin embargo, creo que este no es el único aspecto relevante del procesamiento neuronal. El procesamiento neuronal es la actividad que lleva a cabo un sistema específico de ciertos organismos biológicos, el sistema nervioso. Como tal, parece razonable pensar que la función misma del sistema nervioso, aquello que aporta al comportamiento del organismo del que forma parte, es un elemento crucial para caracterizar la actividad de este sistema. Si esto es así, entonces pasa a ser relevante para evaluar la relación entre el procesamiento neuronal, por un lado, y la computación y la cognición por el otro, determinar la relación entre la función principal del sistema nervioso, el control por retroalimentación negativa, y las nociones de computación y cognición.

La función específica de la actividad neuronal permite distinguir al sistema nervioso de otros sistemas funcionalmente organizados. El sistema nervioso pertenece a la clase de los sistemas de control por retroalimentación negativa junto con los sistemas de auto-piloto, por ejemplo. Piccinini y Bahar (2013) señalan que la tesis de que el sistema nervioso es un sistema de control por retroalimentación no es trivial y tomo mucho trabajo científico establecerla rigurosamente (Cannon 1932). Sin embargo, hoy en día ya no es una tesis controvertida, aunque, como veremos, sí hay debate sobre la naturaleza del control por retroalimentación mismo. Piccinini y Bahar caracterizan al control por retroalimentación nervioso afirmando que el sistema nervioso controla el comportamiento de un organismo en respuesta a los estímulos de que llegan del cuerpo del organismo y de su entorno, incluyendo los efectos del sistema nervioso mismo.

Podemos ofrecer una versión un poco más precisa de esta caracterización de la retroalimentación nerviosa: Para ejercer el control por retroalimentación sobre un sistema S del que forma parte, un sistema S_1 sólo necesita causar un estado B en S o su entorno (controlar el comportamiento de S) y que B sea una respuesta de S_1 a un estado previo A de S , a una conducta previa de S , causada por S_1 mismo. B es un efecto que S_1 produce a causa de otro efecto previo A de S_1 mismo. Dicho en otras palabras, un sistema de retroalimentación es un sistema tal que algunos de los valores de su variable de output causan nuevos valores de la misma variable. Creo que es esto todo lo que necesitamos para que haya por retroalimentación.

Sin embargo, a la luz del debate desarrollado sobre la noción de retroalimentación, es necesario

ofrecer una defensa de esta caracterización. La noción presentada por Wimsatt (1971) es mucho más demandante que la de Piccinini y Bahar (2013). Antes de introducirme en el debate sobre control por retroalimentación, hay un contrañimiento adicional para la retroalimentación neuronal que está ausente en la caracterización de Piccinini y Bahar y que deberíamos señalar. La retroalimentación neuronal es negativa. Muy esquemáticamente, la retroalimentación es negativa ssi cuando la variable de la señal output del sistema controlado se aparta de cierto valor, su valor de equilibrio, el sistema de control la lleva nuevamente a ese valor.

Ha habido un debate substancial sobre la naturaleza del control por retroalimentación. Considero que Wimsatt (1971) ha contribuido con dos ideas que son muy relevantes para pensar la relación entre retroalimentación, computación e información. Wimsatt argumenta que el control por retroalimentación no puede ser caracterizado en términos puramente conductuales, es decir, en términos de cierta correlación entre inputs y outputs específicos de un sistema. Eso quiere decir que para determinar si un sistema lleva a cabo control por retroalimentación uno tiene que determinar, en cierta medida, la estructura interna del sistema. Esto es importante para nuestros propósitos porque, como veremos, que el control por retroalimentación implique computación o procesamiento de información depende exactamente en cuán fuertes son los constreñimientos sobre la estructura interna de un sistema de retroalimentación. Wimsatt (1971) también argumenta que dos condiciones ampliamente aceptadas para caracterizar la estructura interna de un sistema de retroalimentación son insuficientes para que algo cuente como un sistema de este tipo (aunque sí acepta que son necesarias). En la sección 2 argumentaré que la caracterización conductual (o incluso algo más permisivo que ella) es correcta. En la sección 3 mostraré que, sin embargo, el tipo específico de retroalimentación negativa que lleva a cabo el sistema nervioso sí nos dice algo sobre la estructura interna y que los constreñimientos que se consideran típicamente necesarios para la retroalimentación son propios de este tipo más específico de retroalimentación. En la sección 4 mostraré, en primer lugar, que la retroalimentación neuronal no implica computación pero sí implica procesamiento de información. En segundo lugar, mostraré que determinar la relación entre retroalimentación y cognición requiere ahondar en aspectos más complejos de la retroalimentación nerviosa. En particular, argumentaré que es una especie de proceso de meta-retroalimentación neuronal, un proceso que genera mecanismos de retroalimentación neuronal es algo que implica el empleo de conceptos, tal como los he caracterizado en capítulos anteriores.

2. Retroalimentación y “conductismo”

El uso de la noción de retroalimentación en el contexto de la caracterización del comportamiento intencional no es una idea nueva. Si bien aquí mi intención es vincular a la intencionalidad, informacionalmente caracterizada, con el procesamiento neuronal, caracterizado en términos de retroalimentación, ha habido intentos de caracterizar a la intencionalidad misma en términos de retroalimentación. Por ejemplo, Rosenblueth, Wiener y Bigelow (1943, 1950 y 1961), adoptando una perspectiva conductista respecto de la explicación del comportamiento intencional, han propuesto una caracterización de la conducta intencional en términos de retroalimentación negativa. Es interesante reparar en la caracterización del credo conductista que estos autores ofrecen para tener el tipo de concepto de retroalimentación que adoptan. Rosenblueth, Wiener y Bigelow (1943) afirman que

“Dado cualquier objeto, relativamente abstraído de su entorno para ser estudiado, el enfoque conductista consiste en la determinación de los outputs del objeto y de la relación de estos outputs con sus inputs. Por “output” se quiere decir cualquier cambio producido por el objeto en el entorno. Conversamente, por “input” se quiere decir cualquier evento externo al objeto que lo modifica de cualquier manera.

La caracterización anterior de lo que constituye el método conductista omite la estructura específica y la organización intrínseca del objeto. Esta omisión es fundamental porque se basa en la distinción entre el método conductista y el funcional. En un abordaje funcional, a diferencia de lo que ocurre en uno conductista, el objetivo principal es la organización intrínseca del objeto estudiado, su estructura y sus propiedades; las relaciones entre el objeto y el entorno son más bien incidentales.

[...] por conducta se entiende cualquier cambio de una entidad respecto de su entorno, cualquier modificación de un objeto observable externamente puede ser caracterizada como conducta.” (1943, pp. 19-20)

A su vez, los autores caracterizan el comportamiento intencional (o “teleológico”) en términos de retroalimentación negativa. Caracterizan a esta última noción del siguiente modo:

“En un sentido amplio, significa que parte de la energía output del aparato o máquina vuelve a ingresar como input. Un ejemplo es un amplificador eléctrico con retroalimentación. La retroalimentación en este caso es positiva. La fracción del output que vuelve a entrar en el objeto tiene el mismo signo que el input original. La retroalimentación positiva aumenta la señal original, no la corrige. La noción de retroalimentación es también empleada en un sentido más restringido para referirse a objetos cuyo comportamiento es controlado por un margen de error relativo a un objetivo específico. En estos casos la retroalimentación es negativa. Las señales del objetivo son empleadas para restringir la conducta que de otro modo iría más allá del objetivo. Este es el sentido de retroalimentación que emplearemos aquí. Entendemos que un objeto tiene conducta sin retroalimentación cuando las señales del objetivo no lo modifican en el curso de la conducta.” (1943, p. 18)

Uniando la tesis conductista de los autores con su caracterización de la retroalimentación negativa podemos ofrecer una formulación más precisa de esta última noción. Vimos que, según el enfoque de estos autores, la conducta de un objeto, sus diferentes estados observables, puede ser entendida plenamente determinando la relación de sus diferentes conductas con sus inputs, estados del entorno que causan cada conducta, y sus outputs, los estados del entorno que cada conducta causa. A su vez, la conducta de un objeto dado tiene retroalimentación sí y sólo si algunos de sus outputs, los estados del entorno producidos por alguna de sus conductas, se convierten a su vez en inputs, es decir, causan nuevas conductas, nuevos estados en el objeto. Así, parece que podemos caracterizar la retroalimentación según estos autores por medio de una variable x que represente estados de conducta y una constante e que represente un estado del entorno. Así, la retroalimentación ocurre cuando un valor a de x causa e en el entorno y e causa a su vez un nuevo valor b en x .

Es importante notar que esta caracterización de la retroalimentación difiere en algunos aspectos de la noción de Piccinini y Bahar que presenté en la sección anterior. En primer lugar, como Piccinini y Bahar (2013) caracterizan a la retroalimentación como una función del sistema nervioso, toman no sólo a estados del entorno sino a la conducta como el output mismo del proceso de retroalimentación. La retroalimentación neuronal causa conductas a partir de conductas que ha causado previamente. Sin embargo, más allá de esta diferencia, en ambos casos la retroalimentación ocurre cuando hay un loop causal entre diferentes conductas del objeto, cuando una conducta dada da lugar a un proceso causal que culmina con una nueva conducta. Lo que tienen en común estas

caracterizaciones es que prescinden de toda información sobre los procesos internos del objeto responsables de ese loop causal. Si bien en la caracterización de Rosenblueth, Winer y Bigelow hay un estado externo que, al funcionar como output/input, media la relación causal entre las diferentes conductas, no se dice nada sobre los *estados internos* del objeto que median dicha relación.

Como señala Wimsatt (1971), varios autores como Taylor (1950 y 1966), Scheffler (1959 y 1963) y otros han criticado este enfoque sobre la intencionalidad en base a sus aspectos conductistas. Sin embargo, no se han ofrecido críticas a esta caracterización de la noción de retroalimentación. Por el contrario, Wimsatt considera que el problema de la propuesta de Rosenblueth, Weiner y Bigelow (1943) es que la noción misma de retroalimentación no puede caracterizarse en términos puramente conductuales, como una correlación de inputs y outputs externos al objeto, mediados por estados observables en este.

Como vimos en capítulos anteriores, considero que la intencionalidad debe ser explicada en términos de estados internos del sujeto, como creencias y deseos. Considero que este tipo de explicación es viable porque podemos dar una caracterización naturalista de esos estados en términos informacionales. Este enfoque informacional es entonces incompatible con una caracterización conductista de la intencionalidad, que no apela a estados internos. Ahora bien, si, como argumenta Wimsatt (1971), la noción de retroalimentación sí nos dice algo sobre los estados o estructura interna del objeto que la posee, entonces tal vez sí nos diga algo sobre la cognición, entendida como el procesamiento de estados internos. En la última sección de este capítulo consideraré la relación entre estas nociones.

Por el momento nos importa, entonces, el argumento de Wimsatt en contra de la caracterización conductual de la noción de retroalimentación. Wimsatt toma la prueba formal de Kleene (1956) para establecer su tesis. Kleene (1956) mostró que cualquier matriz de input-output (con ciertas restricciones²), puede ser producida por una red de las llamadas “neuronas de McCulloch-Pitts”. Estas “neuronas” propuestas por McCulloch y Pitts (1943) son elementos lógicos generalizado de una lógica bivalente, modelados a partir de una simplificación del comportamiento de las neuronas reales y que,

²Los elementos que median entre input-output deben ser secuencias finitas de elementos discretos con puntos temporales de comienzo definido, tal que cada uno de estos elementos puede ser caracterizado por un conjunto finito de variables, cada una de las cuales sólo puede tomar un número finito de valores.

por medio de sus conexiones con la lógica simbólica puede ser fácilmente manipulado para determinar la conducta de redes interconectadas de estas neuronas. Los resultados de McCulloch y Pitts fueron extendidos por Kleene (1956). Además de la tesis recién mencionada (que cualquier matriz input-output puede ser producida por una red de neuronas de McCulloch y Pitts), Kleene muestra que cualquiera de estas matrices input-output puede ser producida por una red que no contenga lo que Kleene llama “círculos” (p. 10 teorema 1). Wimsatt considera que estos círculos, en la terminología de McCulloch y Pitts, no son otra cosa que lo que llamaríamos un loop de retroalimentación (p. 245). Pero entonces, afirma Wimsatt, si cualquier matriz conductual de input-output puede ser realizada por una red que no contenga loops entonces ninguna matriz de input-output implica que el sistema estudiado tiene un loop de retroalimentación, positiva o negativa. Wimsatt concluye a partir de esto que es imposible definir a la retroalimentación en términos conductistas.

Las redes de neuronas de McCulloch y Pitts son entonces elementos que pueden producir cualquier patrón de input-output, incluyendo los característicos de la retroalimentación. Pero además pueden producir el patrón de input-output propio de la retroalimentación *sin que se de el loop causal propio de la retroalimentación*. Antes de evaluar este argumento es importante notar que refuta algo más fuerte que la noción conductista de retroalimentación. Muestra que una correlación entre un input y un output no basta para determinar que esos dos elementos tienen retroalimentación, *aún si esos elementos no son observables externamente*, como requiere la concepción conductista. Hay loops causales que se dan entre variables de un organismo, no observables externamente. Por ejemplo, el ciclo respiratorio determina el valor de dióxido de carbono en sangre, este valor a su vez es detectado por las áreas del cerebro encargadas de regular la respiración y consiguientemente influyen sobre el ciclo respiratorio. Aquí tenemos loops causales entre variables de un organismo no observables externamente. Si el argumento de Wimsatt funciona, nos mostraría que aún para determinar que las correlaciones entre variables internas como estas constituyen un caso de retroalimentación, la correlación de variables no basta. Esto es importante, porque nos muestra que la noción de retroalimentación criticada es demasiado estrecha, no hay motivo para restringir los casos de retroalimentación a aquellos en los que las variables son observables externamente. Por lo tanto, si logro desestimar el argumento de Wimsatt, considero que la noción defendida no es estrictamente conductista.

Pasemos ahora a considerar el argumento. Sostendré que del hecho de que una red sin círculos puede producir el patrón de input-output de una red con círculos no implica que el patrón input-output

de una red que posea *retroalimentación* puede ser producida por una red que no la posea. Para ver esto, hay dos cosas que tenemos que caracterizar: Las nociones de neurona de McCulloch y Pitts y la de red circular de neuronas. Argumentaré que estas dos nociones nos permiten ver que el argumento de Wimsatt falla.

Según Kleene (1956, sección 3), una neurona de McCulloch y Pitts está constituida por un cuerpo o soma, cuyas fibras nerviosas o axones terminan en un bulbo. Una red de neuronas es una organización de un conjunto finito de neuronas tal que el bulbo de cada neurona es adyacente (y por lo tanto afecta) el cuerpo de no más de una neurona (la misma u otra). Cada bulbo es o bien excitatorio o bien inhibitorio. Las neuronas a las que no afecta ningún bulbo son las neuronas inputs, las demás son neuronas internas de la una red dada. En un momento dado de tiempo, toda neurona está o bien disparando o bien no disparando. Para la neurona input, el que dispare en un momento dado, depende de condiciones externas a la red. Kleene supone que estas neuronas son afectadas por un órgano sensorial que causa que cuando se dan ciertas condiciones en el entorno, la neurona input dispare. Para una neurona interna, la condición de que dispare en un momento dado es que un cierto número n (el umbral de esa neurona interior) de bulbos excitatorios y ninguno de los bulbos inhibitorios que la afectan pertenezcan a neuronas que dispararon en un momento anterior.

Por su parte, un círculo de longitud c en una red neuronal es, según Kleene, un conjunto de neuronas $N_1 \dots N_c$, (c igual o mayor a 1) tal que el bulbo de N_i afecta a N_{+i} , para todo i ($i= 1, \dots, c-1$) y el bulbo de N_c afecta a N_1 . Es decir, un círculo es una red de neuronas tal que la neurona “input” está conectada por medio su bulbo, o bien con si misma o bien con neuronas cuyos bulbos afectan esa misma neurona “input”. Me refiero al input por medio de comillas ya que, al ser afectaba por un bulbo, dicha neurona no es un input en el sentido de Kleene. Ahora podemos caracterizar más precisamente la tesis que Wimsatt toma de Kleene.

El input y el output que un círculo relaciona son, por definición, neuronas N_1 y N_c . La tesis de Kleene no puede ser, sin embargo, que para todo círculo que contenga a N_1 como input y a N_c como output, hay una red no circular que tiene el mismo input y output. Pues esto no necesita prueba. Para generar esta segunda red basta meramente con que el bulbo de N_c no esté conectado con N_1 . Lo que nos interesa del círculo es que genera loop causal, es decir, una retroalimentación entre disparos (o no disparos) de N_1 y N_c . La idea de Kleene que a Wimsatt le interesa es que esta correlación puede ser generada por una red que no constituya un círculo. Es decir, la correlación entre los disparos de N_1 y N_c

puede ser reproducida por una red en la que N_1 y N_c no están conectados por medio del bulbo de N_c . Ahora bien, no sólo estoy dispuesto a aceptar dicha conclusión, sino que considero que este fenómeno se da efectivamente con frecuencia y es lo que consideramos normalmente como la retroalimentación que produce el sistema nervioso.

Como vimos, una neurona genuinamente input se dispara, según Kleene, a causa de la presencia de un estímulo externo a la red y no por la actividad de otra neurona con cuyo bulbo está conectada. A su vez, Kleene asume que es la actividad de algún conjunto de neuronas la responsable de que el organismo lleve a cabo determinada acción o de que tenga cierta conducta (sección 1). Supongamos que un disparo de N_1 , el input de la red que constituía un círculo, causa un disparo de la neurona N_c , el output de la misma red. Supongamos además que el disparo de N_c causa ahora, en lugar del disparo de N_1 , una acción A por parte del organismo que posee dichas neuronas. Supongamos que A genera un hecho H en el entorno, que H causa un estímulo E y, por último, que E causa que N_1 se dispare nuevamente. Aquí tendremos un loop causal entre N_1 y N_c sin que haya un círculo en el sentido de Kleene, sin que el bulbo de N_c esté conectado con N_1 . Esto quiere decir que el hecho de que pueda haber una correlación entre los disparos de N_1 y N_c sin que haya un círculo entre ellos no implica que pueda haber una correlación entre ellos sin que haya un loop causal. Por lo tanto, el argumento de Wimsatt no muestra que necesitamos determinar algo más que la correlación entre un input y un output para mostrar que hay un loop causal entre ellos, sino sólo que necesitamos algo más que la correlación para determinar que hay un círculo.

De este modo, perdemos la motivación que propone Wimsatt para adoptar una noción más específica que la noción conductista de retroalimentación. Ahora bien, dije antes que la noción conductista era aún muy restringida, nuestra noción intuitiva de retroalimentación se puede aplicar a muchas variables que no son observables. Pero creo que aún permitiendo que se extienda de este modo, la noción es aún muy restrictiva. Tal como la caracterizaban Rosenblueth, Wiener y Bigelow, la retroalimentación es un patrón circular de relaciones causales que se da entre conductas de un objeto. Como vimos, en la caracterización de estos autores, el loop está mediado por un estado del entorno e que funciona como output/input. Una conducta causa un estado externo que produce una nueva conducta. Sin embargo, hay una descripción más simple de este loop causal, y no es otro que el descrito en la caracterización que Piccinini y Bahar ofrecen de la retroalimentación. Como vimos, estos autores caracterizan a la retroalimentación (o mas bien, su caracterización puede modelarse como) haciendo referencia a una sola variable que representa conductas/estados del entorno. Podemos

decir que un sistema tiene retroalimentación cuando un valor de su variable x de output se convierte en el input de un proceso que culmina con un nuevo valor de la misma variable. Esta es una caracterización aún más simple que la que podíamos ofrecer según el modelo conductista considerado.

En la próxima sección argumentaré que si bien la noción general de retroalimentación no requiere más que la correlación causal de diferentes valores de una misma variable, la idea de retroalimentación neuronal, la retroalimentación llevada a cabo específicamente por el sistema nervioso sí requiere algo más que esto.

3. Retroalimentación neuronal

Como vimos, la noción de retroalimentación requiere muy poco para que un sistema la posea. De acuerdo con mi defensa de Piccinini y Bahar (2013) podemos caracterizarla en base a una sola variable y decir que un sistema posee retroalimentación si y sólo si los valores de su variable output causan nuevos valores de la misma variable. Para el caso de la retroalimentación negativa, el tipo que posee el sistema nervioso, podemos decir que un sistema tiene retroalimentación negativa si y sólo si cuando su variable output (que representa diferentes conductas) tiene cierto valor P diferente de su valor O (la conducta O que conduce al resultado específico deseado) esto causa que el output vuelva a ser O .

Considero que cuando decimos que el sistema nervioso es un sistema de retroalimentación negativa no queremos decir que se da un loop causal entre valores de alguna variable suya sino principalmente que *produce* un loop causal en una variable *externa* al sistema. Esto es por lo siguiente. La retroalimentación nerviosa es la función de un sistema *de un organismo dado*, es decir, es algo que contribuye al funcionamiento de dicho organismo. Por lo tanto, parte de lo que determina la función del sistema tiene que ser presumiblemente alguna variable externa al sistema mismo, una variable de estados del organismo del que forma parte y/o su entorno. En este sentido quisiera distinguir entre decir que un sistema tiene retroalimentación y decir que un sistema *produce* retroalimentación. Considero que cuando decimos que el sistema nervioso es un sistema de retroalimentación tenemos principalmente el segundo sentido en mente.

Que el sistema nervioso produce un loop de retroalimentación entre las conductas que causa no

quiere decir otra cosa que el sistema nervioso es un mediador causal entre dichas conductas. Una conducta *A* del organismo *produce* un estado *N* en el sistema nervioso que, a su vez, causa una nueva conducta *B* en el organismo. Ahora bien, como señala la caracterización de Piccinini y Bahar, la primera conducta (la que vuelve a entrar como input) es también un efecto del sistema nervioso. Suponiendo que tenemos un estado *M* del sistema nervioso, responsable de producir *A*, se ve que al producir un loop causal entre conductas *A* y *B*, se genera un loop causal entre los estados internos *M* y *N* del sistema nervioso. *M* causa *A*, pero *A* modifica el estado del sistema nervioso de *M* a *N* (el estado que modifica a su vez la conducta). Al mediar la relación causal entre las conductas que produce, se genera un loop causal interno al sistema nervioso. Así, *producir* retroalimentación en una variable externa, en el sentido en que el sistema nervioso lo hace, implica *tener* retroalimentación en una variable interna.

Ahora bien, aunque hay una restricción para la estructura interna de los sistemas que producen retroalimentación, esta restricción difiere del tipo de restricción que Wimsatt cree que estos sistemas tienen. Pues el argumento de Wimsatt intentaba establecer justamente que el patrón input-output de un sistema de retroalimentación no basta para determinar que es un sistema de este tipo sino que hace falta algún aspecto específico de su estructura interna *que no es implicado por dicho patrón*. Por el contrario, argumenté aquí que de la noción de sistema que produce retroalimentación *se sigue* algo específico sobre la estructura interna. Hay una relación *conceptual* entre el patrón input output y la estructura interna de los sistemas que producen retroalimentación.

Otro punto importante es que el aspecto de la estructura interna que es implicado por la producción de retroalimentación es otro caso de retroalimentación. Para que se de el loop causal entre dos valores de la variable output, los valores de alguna variable interna tienen que mediar ese loop por medio de otro loop de retroalimentación. Producir retroalimentación implica tener retroalimentación. Sin embargo tener retroalimentación no implica producirla. Las conductas *A* y *B* de un objeto podrían tener una relación causal que no esté mediada por otros estados.

Además de este vínculo conceptual entre retroalimentación y estructura interna considero que que la retroalimentación nerviosa así caracterizada tiene vínculos importantes con condiciones que se toman frecuentemente como necesarias de los mecanismo de retroalimentación. Principalmente, esta noción de retroalimentación implica las condiciones de equifinalidad y de compensación. Esto tiene una consecuencia importante para el enfoque que estoy presentando sobre la retroalimentación.

Wimsatt argumentó que como el patrón input-output no permite determinar la retroalimentación debemos buscar condiciones diferentes, restricciones para la estructura interna de los sistemas de retroalimentación. Así considera las condiciones de equifinalidad y compensación como condiciones adicionales y propone además una propia. Por el contrario, mostraré que la idea general de un sistema de retroalimentación *implica conceptualmente* las condiciones de equifinalidad y compensación. Mi enfoque muestra que las condiciones de equifinalidad y compensación están ya contenidas en la noción de retroalimentación, no son condiciones adicionales. Pasemos a ver en qué consisten estas condiciones.

Wimsatt considera principalmente dos condiciones que han sido ampliamente aceptadas para el control por retroalimentación. Una condición es que un sistema sea capaz de alcanzar un mismo conjunto de valores de equilibrio para sus variables de estado, desde un rango arbitrariamente amplio de diferentes valores iniciales de las mismas variables de estado (esta es la característica principal del análisis de Braithwaite 1953). Von Bertalanffy llamó a este fenómeno ‘equifinalidad’ y tanto él como Kacse (1957) ofrecieron análisis formales de este fenómeno para sistemas abiertos de tipo general que Wimsatt afirma que no son sistemas de retroalimentación. Así, aunque esta característica sea necesaria para la retroalimentación, no es suficiente.

La segunda condición que considera es que los cambios en una o más variables de estado serán parcialmente o completamente compensados por cambios en una o más variables de estado de modo tal que la variable original perturbada recuperará parcial o totalmente su valor normal de equilibrio. Este fenómeno ha sido llamado por Kacser (1957) ‘amortiguación’ (“buffering”) una expresión que es una mejor guía a su significado que las empleadas en el análisis mencionado antes. La equifinalidad describe el efecto de la amortiguación o compensación. Esta condición de compensación describe el mecanismo por el que la equifinalidad de los estados de no equilibrio es lograda: algunas de las variables toman la carga de las variables estresadas y les permiten volver parcial o totalmente a sus valores normales.

Considero que la noción general de retroalimentación (la noción de *tener retroalimentación*), tal como la caractericé, en base a una sola variable, implica equifinalidad pero no compensación. Como vimos, tener retroalimentación negativa requiere que cuando el valor del output es un valor V diferente del valor de equilibrio E esto cause que la variable vuelva a E . La noción general de retroalimentación implica entonces el fenómeno de equifinalidad. Por su parte, la compensación ocurre cuando el cambio

de valor de la variable output de V a E ocurre por medio del cambio de valor de otra variable (la que toma la carga de la variable output). Como no es conceptualmente necesario tener más de una variable para que ocurra el loop causal típico de la retroalimentación, parece que la compensación no es necesaria para ella. Sin embargo, creo que la propiedad de producir retroalimentación, sí implica compensación. Vimos que en este tipo de retroalimentación, cuando un sistema tiene la función de producir retroalimentación en una variable externa al mismo, tenemos al menos dos variables. Una variable input/output externa al sistema y otra variable para estados del sistema. En el caso de la retroalimentación negativa, cierto valor W en la variable interna produce un valor V diferente a E en la variable input/output y a su vez, V produce un nuevo valor Y en la variable interna que causa que el valor de la variable input/output sea E . El cambio de V a E en la variable input/output es causado por el cambio de valor de la variable interna que, en este sentido, *toma la carga de la variable externa*. De este modo, la noción de producir retroalimentación, al implicar las dos condiciones consideradas comúnmente necesarias para la retroalimentación, muestra que estas no son condiciones adicionales, no contenidas en la noción de retroalimentación.

He ofrecido en esta sección una caracterización de la retroalimentación nerviosa. En primer lugar he sostenido que el sistema nervioso puede ser caracterizado como un sistema que no sólo tiene retroalimentación sino que produce retroalimentación en un sistema más amplio del que forma parte. En segundo lugar, he mostrado que la retroalimentación nerviosa así caracterizada implica dos condiciones consideradas frecuentemente como necesarias para la retroalimentación.

4. Retroalimentación, computación, procesamiento de información y cognición

Como vimos en la sección 2, las críticas más salientes en contra de definir el comportamiento intencional en términos de retroalimentación están ligados al carácter conductista de esta noción. Por su parte, vimos que Wimsatt cuestiona la caracterización conductista de la retroalimentación. Argumenté que la noción adecuada es incluso una más simple que la conductista. Sin embargo, mostré en la sección anterior que la retroalimentación nerviosa sí presenta algunas características adicionales. En particular presenta alguna restricción para la estructura interna del sistema, la responsable del loop causal entre valores de la variable input/output. Ahora bien, es interesante preguntarnos si esta noción más estrecha de retroalimentación nerviosa implica algo sobre la intencionalidad tal como la hemos

entendido en los capítulos anteriores. Si bien una noción conductista de retroalimentación está desligada de la cognición entendida en términos de contenidos representacionales y sus vehículos computacionales, la noción de retroalimentación nerviosa podría tener alguna relación con estas nociones. En esta sección argumentaré que la noción de retroalimentación nerviosa implica procesamiento de información pero no computación. Además mostraré que a pesar de que implica procesamiento de información no implica cognición. Si los vehículos de la cognición son los conceptos, tal como los caractericé, entonces implican algo más específico que el procesamiento de información. Argumentaré, sin embargo, que aspectos peculiares de la retroalimentación neuronal en el sistema nervioso humano sí implican cognición. Si bien rechazo la caracterización de la intencionalidad en términos de retroalimentación, ya que considero que las acciones intencionales deben explicarse en virtud de contenidos conceptuales, si aceptamos una concepción informacional de la cognición, el hecho de que el sistema nervioso produzca retroalimentación de cierto tipo implica que es un sistema cognitivo en un sentido no conductista, es un sistema cuyo comportamiento puede ser explicado, en virtud de los contenidos de las representaciones que manipula.

Empecemos considerando la relación entre retroalimentación nerviosa y computación. Vimos que el sistema nervioso tiene como función producir retroalimentación en una variable output de estados externos al mismo, una variable y de estados del organismo del que forma parte o su entorno. Vimos que esto no requiere más que una variable interna con dos estados que causen los valores de las variables externas y que los conecten causalmente entre sí. Además hemos visto que como la retroalimentación nerviosa es negativa, uno de los valores de la variable externa debe representar un estado de equilibrio mientras que otros valores deben representar un estado de no equilibrio. El estado de no equilibrio de la variable de conducta/entorno causa un estado en la variable interna que, a su vez, causa un estado de equilibrio en la variable de entorno/conducta.

Ahora bien, es sencillo notar que estas restricciones no implican que los vehículos de la retroalimentación neuronal deban ser computacionales en ningún sentido. Las restricciones para los valores de estas variables son puramente extrínsecas. Lo único que se necesita es que diferentes valores de estas variables se encuentren en determinadas relaciones causales con valores de otras variables. Al igual que vimos para el caso de los vehículos del procesamiento de la información, cuando las restricciones son extrínsecas en este sentido, no se sigue nada respecto de las propiedades intrínsecas de los vehículos relevantes. En particular, queda abierta la posibilidad de que las variables mismas sean continuas o discretas. El control neuronal por retroalimentación no implica computación porque sus

restricciones dejan abierta la posibilidad de que los aspectos funcionalmente relevantes de sus variables sean muy específicos o concretos. Por lo tanto, la retroalimentación neuronal no implica relevancia funcional restringida y por lo tanto tampoco computación.

Podemos ahora considerar la relación entre el control por retroalimentación y el procesamiento de información. La relación entre retroalimentación e información está muy poco trabajada por Piccinini y Bahar (2013). Ellos sólo afirman que la complejidad de la retroalimentación llevada a cabo por el sistema nervioso hace necesario que el sistema procese variables internas que portan información sobre variables externas:

“El único modo eficiente de ejercer funciones de control suficientemente complejas para el sistema nervioso es poseer y procesar variables internas que están confiablemente correlacionadas con variables externas relevantes. Estas variables internas deben ser procesadas de acuerdo con la información natural que portan” (2013, p. 4).

Piccinini y Bahar no argumentan, sin embargo, por qué la complejidad de la retroalimentación neuronal hace necesario el procesamiento de información ni determinan si esta necesidad es conceptual, empírica o de otro tipo. En lo que sigue, argumentaré que incluso un modelo de retroalimentación negativa neuronal mucho más débil que la neuronal implica *conceptualmente* procesamiento de información.

Vimos que el modelo mínimo para representar a un sistema que produce retroalimentación involucra al menos dos variables: una variable y para estados externos al sistema y una variable x de estados internos al sistema. La variable x tiene al menos dos estados, un estado de equilibrio E y un estado de no equilibrio $no-E$. Cuando algún valor A de x produce $no-E$ en y , $no-E$ produce en x un valor B que produce E en y . Asumiendo que las correlaciones causales son confiables, se sigue entonces que sólo B portará la información de que y tiene un valor de no equilibrio y sólo B producirá el valor de equilibrio. Vimos en el capítulo anterior que el procesamiento de información requiere una correlación de información y poderes causales. Un sistema que procesa información responde de cierto modo específico a un estado que porta cierta información I pero no a otros estados (estados que portan otra información o que no portan información en absoluto). De esto se sigue que la retroalimentación negativa implica procesamiento de información.

¿Qué implica esto específicamente respecto de la relación entre procesamiento neuronal y cognición? Si aceptamos la caracterización informacional que he ofrecido de los vehículos del procesamiento cognitivo, de los conceptos, entonces se sigue que retroalimentación neuronal no implica cognición. Como vimos en capítulos anteriores, el contenido conceptual de una estructura *C* está constituido por dos tipos de contenido informacional. La información que *C* mismo tiene la función de portar, y la información que tienen la función de portar las estructuras que figuran en las inferencias por default de *C*. Esto quiere decir que un concepto requiere de *diversas* estructuras que porten información y que estén causalmente relacionadas entre sí. Por el contrario, el modelo mínimo de retroalimentación propuesto no implica más que un estado informacional.

Ahora bien, para determinar la relación entre la cognición tal como la entiendo y la retroalimentación nerviosa podemos hacer el camino inverso. Podemos preguntarnos en qué aspecto de la retroalimentación podría influir la cognición tal como la entiendo, es decir, la capacidad de relacionar causalmente diferentes estados informacionales (presumiblemente mediante enlaces de memoria). Creo que hay un aspecto crucial para la retroalimentación en lo que puede contribuir esta capacidad. Si estoy en lo cierto, este es el vínculo que buscábamos entre la cognición y el procesamiento neuronal.

La capacidad de enlazar diferentes piezas de información nos permite, entre otras cosas, adquirir conocimiento sobre relaciones causales entre tipos de eventos. Por ejemplo, podemos enlazar la representación visual de un alimento con la representación gustativa del mismo, aprender que los objetos que tienen dicha apariencia producen cierta sensación. Más específicamente, también podemos aprender sobre las relaciones causales entre nuestras conductas y sus resultados. Podemos conectar la representación propioceptiva de una conducta corporal *C* con la representación del resultado específico *R* que produce. En particular, podemos aprender, respecto de cualquier resultado específico, qué conductas no conducen a él, qué conductas conducen a él y, entre estas últimas, cuáles conducen más efectivamente al mismo resultado. Es claro el efecto que esto puede tener sobre nuestra conducta. En primer lugar, podemos aprender a buscar nuestros objetivos con más precisión, restringir más el rango de conductas adecuadas para obtener cierto resultado. Es decir, dado un objetivo que tengamos, podemos restringir, *refinar nuestros valores de equilibrio* para obtenerlo. Esto quiere decir que podemos refinar nuestros loops de retroalimentación. En segundo lugar, podemos formar loops de retroalimentación totalmente nuevos al aprender, respecto de un resultado totalmente nuevo de nuestra conducta, qué conductas específicas conducen a él y cuáles nos alejan. Es decir, podemos generar

respecto de nuevos objetivos, valores de equilibrio y valores de no equilibrio de la variable de nuestra conducta.

Así, la capacidad de relacionar diferentes piezas de información nos permite producir *tipos* de loops de retroalimentación nuevos. Dije antes que el sistema nervioso es un sistema que produce retroalimentación. Con esto quería decir que dado un loop de cierto tipo entre conductas específicas *A* y *B*, el sistema nervioso funciona como un mediador causal para los casos de retroalimentación de ese tipo específico. La capacidad de relacionar información le permite al sistema nervioso producir retroalimentación a otro nivel. Puede producir nuevos tipos de retroalimentación, es decir, loops nuevos que involucren nuevos estados de equilibrio y no equilibrio respecto de un objetivo ya dado o incluso de objetivos completamente nuevos. Así el sistema nervioso no sólo interviene causalmente en los procesos de retroalimentación específicos, sino que es capaz de producir nuevos tipos de procesos de retroalimentación.

5. Conclusión

En este capítulo he intentado restablecer (después de quebrar el vínculo entre cognición y computación en el capítulo pasado) el vínculo conceptual entre el procesamiento neuronal y la cognición. Para ello me centré en la caracterización del procesamiento neuronal en base a la función del sistema que lo lleva a cabo, la retroalimentación neuronal del sistema nervioso.

Mostré que Piccinini y Bahar mismos ofrecen una caracterización muy simple de esta noción y que este tipo de caracterización ha sido puesta en discusión principalmente por Wimsatt. Wimsatt argumentó que el hecho de que una red de neuronas pueda tener el patrón input/output de un sistema de retroalimentación sin tener círculos implica que dicho patrón no permite determinar que la red tenga retroalimentación. Argumenté que esta implicación no se sigue porque la ausencia de círculos en el sentido de Kleene no implica la ausencia de retroalimentación. A partir de esto defendí una noción intuitiva de retroalimentación que no implica nada respecto de la estructura interna que hace que se vinculen causalmente los valores de la variable input/output. Argumenté que, sin embargo, cuando decimos que el sistema nervioso es un sistema de retroalimentación queremos decir que produce retroalimentación en una variable externa al mismo, que media en la relación causal entre los valores

de esta variable externa y que esto sí implica algo sobre la estructura interna del sistema nervioso. En particular, implica que hay al menos una variable interna que también tiene retroalimentación. Vimos que esta noción más restringida de retroalimentación implica conceptualmente las dos condiciones que comúnmente se le adjudican a los sistemas de retroalimentación: equifinalidad y compensación. Con esto finalizamos la discusión sobre la noción misma de retroalimentación.

En la última sección intenté relacionar esta noción con las de computación, información y cognición. Mostré que la noción de retroalimentación neuronal no implica computación pero sí procesamiento de información. Sin embargo, mostré también que la retroalimentación neuronal no implica cognición. Puede haber retroalimentación neuronal sin cognición, entendida como la manipulación de conceptos, en los términos en los que los caractericé en capítulos anteriores. Me pregunté entonces en qué aspectos de la retroalimentación neuronal podría influir la cognición. Si esta pregunta tenía alguna respuesta, esto implicaría al menos que, en los organismos capaces de cognición, el tipo específico de retroalimentación que poseen implica cognición. Mostré que los conceptos, constituidos en parte por relaciones entre piezas de información, aportan una capacidad que puede afectar notablemente la retroalimentación neuronal. Pueden aportar el conocimiento de las relaciones causales entre conductas y resultados de dichas conductas. Argumenté que esta es una herramienta que permite generar nuevos loops de retroalimentación o refinar los que ya posee el organismo. La cognición permite que la retroalimentación nerviosa se desarrolle, tanto en nivel de precisión como en el rango de objetivos que puede abarcar.

Con esto pretendo finalizar la tarea de elucidación de las nociones involucradas en los diferentes niveles del procesamiento cognitivo y la determinación de sus relaciones. En el próximo capítulo intentaré intervenir en el debate contemporáneo en torno a la noción de realizabilidad, proponer una elucidación de dicha noción y determinar, en base a la tarea realizada en los capítulos previos, si la cognición, la computación y el procesamiento neuronal guardan una relación de realización y de qué tipo.

Capítulo 10:

Realización

1. Relaciones conceptuales y relaciones metafísicas

En los tres capítulos previos de esta segunda parte he intentado elucidar las nociones centrales vinculadas con la noción de cognición, según el análisis que llevé adelante en la primera parte: Las nociones de información, computación y procesamiento neuronal. Además, en cada uno de estos capítulos he establecido qué relaciones conceptuales se dan y qué relaciones no se dan entre estos conceptos. En particular, he argumentado que la noción de información no implica a la de computación (ni viceversa), que la noción de vehículo neuronal implica computación pero no información, que retroalimentación neuronal implica procesamiento de información y que cierta forma compleja de retroalimentación implica cognición. Ahora bien, como señalé al comienzo de esta segunda parte, determinar relaciones conceptuales entre estas nociones puede tener dos usos. En primer lugar puede aportar premisas a argumentos a favor de que el sistema nervioso tiene ciertas propiedades. Hemos visto que, por ejemplo, Piccinini y Bahar (2013) emplean estas relaciones para determinar que el sistema nervioso es un sistema computacional. Por otro lado, he mencionado que estas relaciones pueden emplearse para determinar el tipo de relación metafísica que tienen las propiedades a las que estas nociones refieren. Hemos visto, por ejemplo, que el vínculo conceptual entre computación e información lleva a Fodor (1994) a reformular la noción tradicional de realización. También vimos que el tipo de relación conceptual que guarda la noción de vehículo computacional y la de vehículo neuronal parece implicar que estos dos tipos de vehículos no pertenecen a niveles diferentes, que los vehículos neuronales no son más que un tipo específico de vehículo computacional.

Sin embargo, considero que toda consecuencia que podamos extraer de relaciones conceptuales

entre nociones sobre las relaciones metafísicas entre las propiedades a las que refieren, es sólo provisional hasta que no tengamos una idea clara de qué significan las nociones metafísicas relevantes. En este capítulo quiero introducirme en discusiones recientes en torno a la noción de realización con el propósito de fundar tesis metafísicas sobre las propiedades consideradas en los capítulos anteriores.

En la sección 2 presentaré los principales enfoques en torno a la realización y un breve mapa de la discusión reciente en torno a ellos. Como veremos, se han discutido recientemente dos enfoques centrales en torno a la realización, el enfoque dimensionado y el enfoque plano. Argumentaré, entre las secciones 3 y 4, que si bien el enfoque plano es más adecuado en varios aspectos que el dimensionado, debe incorporar algunos aspectos del enfoque dimensionado para resolver determinados problemas propios. En la sección 5 mostraré que, como resultado de este enfoque híbrido, resultará una noción de realización que se aplica tanto a propiedades que están conceptualmente relacionadas (como a la información y la retroalimentación) como a propiedades que no lo están (como a la información y la computación).

2. La realización dimensionada y la realización plana³

Hay dos fuentes de donde surge la noción de realización y la tesis de la realizabilidad múltiple. Una primer fuente es un trabajo de Putnam. Un argumento propuesto por Putnam (1967) puso en duda un supuesto hasta ese entonces ampliamente aceptado sobre la relación entre los estados mentales y los estados físicos: la tesis de la identidad psicofísica. Esta no es otra que la tesis de que cada tipo de estado o propiedad psicológica es idéntico a un tipo de estado o propiedad física del cerebro o el sistema nervioso (Smart 1959). El argumento de Putnam parte de considerar la implausibilidad de que obtengamos la evidencia que necesitaríamos para confirmar un enunciado de identidad psico-física. Por ejemplo, el enunciado que afirma la identidad entre la propiedad de tener dolor y cierto tipo de propiedad física. El teórico de la identidad debe encontrar un único tipo de estado físico-químico que

³La presentación de las fuentes de la tesis de realizabilidad y las posiciones contemporáneas en torno a esta noción están basadas principalmente en una de las reconstrucciones del debate más completas que he encontrado en lengua castellana y uno de los abordajes críticos más agudos en general. El trabajo de Berberis (2014) es además, como veremos, uno de los focos centrales de la discusión que se desarrolla en en las próximas dos secciones.

sea común al cerebro de, al menos, los mamíferos, de los reptiles, de los moluscos, y de todas las especies terrestres capaces de sentir dolor. Como señala Putnam, no se trata de que sea físicamente imposible que, debido a la convergencia de procesos independientes de evolución, esto ocurra, de que exista un estado físico-químico común a todas esas especies. Sin embargo, es altamente improbable que así sea. Lo que sí resulta más probable es que el estado cerebral que, en algún sentido, constituye un mismo tipo de estado psicológico como *dolor*, sea diferente en distintas especies. Por lo tanto, el estado de *dolor* no puede ser idéntico a ninguno de estos tipos de estado cerebral. De esto se concluye que las propiedades psicológicas no son idénticas a propiedades físicas sino a roles funcionales que pueden ser ocupados por estados cerebrales de distinto tipo.

Una segunda fuente de la tesis sobre la realizabilidad es un trabajo de Fodor y Block (1972) que intenta complementar el trabajo de Putnam (1972). Fodor y Block ofrecen principalmente tres fuentes de evidencia a favor de la tesis de la realizabilidad múltiple de lo mental. En primer lugar, los autores señalan la evidencia proveniente del fenómeno de la plasticidad neuronal. Diferentes estructuras neuronales en distintos sujetos o incluso en diferentes momentos en un mismo sujeto, pueden llevar a cabo una misma función psicológica. En segundo lugar, Block y Fodor argumentan que si aceptamos la posibilidad teórica de la inteligencia artificial, se sigue que los estados mentales pueden realizarse en estados físicos no cerebrales. La tercera línea de argumentación parte del fenómeno de la convergencia evolutiva. Es plausible que la selección natural seleccione rasgos por su función psicológica. Si esto es así, entonces es plausible pensar que procesos evolutivos diferentes, que afectan a especies remotas con estados fisiológicos relevantemente diferentes, hayan adquirido una misma propiedad psicológica como resultado de las mismas presiones ambientales. Estas especies tendrían una misma propiedad psicológica realizada en estados fisiológicos diferentes.

El resultado de estos argumentos es, entonces, una tesis peculiar sobre la relación entre las propiedades psicológicas (y muchos otros estados o propiedades no básicas) y las propiedades físicas, un tipo peculiar de fisicalismo. Cuando, en un individuo dado, una propiedad psicológica tiene determinada propiedad neuronal como “base” diremos que dicha propiedad física *no es otra cosa* que la propiedad neuronal en cuestión, no es algo más allá de ella. Sin embargo, no podemos decir que la propiedad psicológica es *idéntica* a la propiedad física. Pues la identidad es una relación de uno a uno, mientras que las propiedades psicológicas pueden ser realizadas en diferentes propiedades físicas. Esta es la peculiaridad de la noción de realización.

Ahora bien, hasta recientemente, no ha habido una tarea de elucidación exhaustiva de esta noción. Hace no más de una década (Barberis 2014), varios filósofos de la mente y de las ciencias cognitivas llevaron a cabo discusión extensa sobre el estatus metafísico de la realización física (Shapiro 2000, 2004; Gillett 2002, 2003, 2007; Polger 2004; Polger y Shapiro 2008; Aizawa y Gillett 2009a, b; Endicott 2011). De un modo un poco esquemático, podemos decir que hay dos enfoques confrontados en torno a la noción de realización. Por un lado, está la concepción “estándar”, “plana” o “intranivel” de la realización. El enfoque plano se atribuye especialmente a Putnam (1967), ocasionalmente a Kim (1998) y Shoemaker (2007), y es reivindicado explícitamente por Shapiro y Polger (2008). Por el otro lado, tenemos el enfoque “dimensionado” o “internivel” de la realización. El enfoque dimensionado es propuesto y defendido por Gillett y Aizawa (Gillett 2002, 2003; Aizawa y Gillett 2009a, b). Estos autores ocasionalmente se lo atribuyen a Block y Fodor (1972) y Fodor (1974). La diferencia intuitiva entre estas perspectivas teóricas es que, mientras el enfoque plano entiende la realización como una relación de dependencia entre un rol funcional y el objeto físico que lo ocupa, el enfoque dimensionado entiende la realización como una relación de dependencia entre una propiedad de un objeto complejo, por un lado, y las propiedades y relaciones de sus partes componentes, por el otro (Barberis 2014).

Antes de mostrar las diferencias específicas entre estos enfoques, Barberis (2014) explicita el núcleo duro de compromisos metafísicos acerca de la realización que ambos enfoques comparten. Consideremos brevemente cada una de ellas. En primer lugar, Barberis afirma que se acepta comúnmente que la realización es un tipo de relación de determinación. Si B es un realizador de A, entonces B determina (de alguna manera) a A, o es metafísicamente suficiente para A. También se puede plantear en términos de dependencia: si B es un realizador de A, entonces A depende de B para su existencia. (Shapiro 2004, p. 36). En la sección 5, argumentaré que, sin embargo, este supuesto no debe ser aceptado por ningún enfoque sobre la realización, que no podemos aceptar, para todo caso de realización que el realizador es metafísicamente *suficiente* para que se de lo realizado, es decir, que no es metafísicamente posible que se de el realizador sin que se de el realizado.

El segundo supuesto que señala Barberis es el de que la realización es una relación asimétrica: si B realiza A, entonces A no puede realizar B. En tercero es que la realización es irreflexiva: ninguna propiedad (o instancia de una propiedad) se realiza a sí misma. En cuarto lugar, el tipo de determinación del realizador sobre lo realizado debe distinguirse netamente de la determinación causal o causación. En particular, se supone que la realización, a diferencia de la causación, no vincula entidades completamente diferentes desde el punto de vista metafísico. Vimos que hay un sentido en el

que la realización es semejante a la noción de identidad: la propiedad realizada no es nada “por encima” del realizador (Craver y Bechtel 2007, Craver 2007). Este supuesto es de suma importancia porque de él depende que la tesis de la realizabilidad múltiple sea compatible con el fisicalismo. Aunque un tipo de propiedad se realice en diferentes propiedades físicas, cada una de sus instancias no es algo diferente de la propiedad física en la que se realiza. Como veremos más adelante, Barberis argumenta, de modo muy sólido, que el enfoque dimensionado es incompatible con este supuesto. Este es, a mi parecer, uno de los problemas centrales del enfoque dimensionado, uno que hace su abandono obligado.

El último supuesto común que señala Barberis es que cualquier análisis de la relación de realización debe ser compatible con el fenómeno de la múltiple realizabilidad: una misma clase de orden superior debe poder estar realizada por instancias de distintas clases de orden inferior. Explicitados estos acuerdos iniciales, podemos pasar ahora a caracterizar cada uno de los enfoques.

Según el enfoque plano, una instancia de una propiedad F realiza una propiedad funcional G sólo si la instancia de F “ocupa” el rol funcional que G especifica. De un modo más preciso, Endicott (2011) propone la siguiente caracterización de la noción plana de realización:

Realización plana: Una instancia de una propiedad F realiza una propiedad G si y sólo si (i) existe algún objeto x que instancia F y que instancia G (ii) G es la propiedad (de segundo orden) de tener una propiedad que ocupa el rol funcional R; y (iii) F ocupa el rol R.

Gillett, uno de los defensores principales del enfoque dimensionado, señala dos compromisos del teórico del enfoque plano, presupuestos en esta caracterización, que lo distinguen del dimensionado. En primer lugar, el teórico plano acepta que una instancia de una propiedad F realiza una instancia de una propiedad G sólo si F y G están instanciadas en el mismo individuo x . En segundo lugar, el enfoque plano acepta que los poderes causales que son individuativos de una instancia de la propiedad realizada G coinciden, o son un subconjunto de los poderes causales que son individuativos de la propiedad realizadora F (cf. Shoemaker 2007). Según Gillett el enfoque dimensionado no se compromete con ninguno de estos supuestos. La propiedad realizada y la realizadora pueden estar instanciadas en distintos particulares y no necesitan compartir ninguno de los poderes causales que las

individúan. El enfoque dimensionado de la realización que proponen Gillett y Aizawa (2009a) puede ser caracterizado de este modo:

Realización dimensionada: Las instancias de las propiedades (o relaciones) F_1 - F_n realizan una instancia de una propiedad G , en un individuo s bajo las condiciones $\$$ si y sólo si, en las condiciones $\$,$ s tiene los poderes que son individuativos de una instancia de G en virtud de los poderes conjuntamente contribuidos por las instancias de F_1 - F_n a s , o a los constituyentes de s , pero no viceversa.

De acuerdo con este enfoque, una instancia de una propiedad en un objeto está generalmente realizada por una colección de instancias de otras propiedades *de sus partes componentes o constituyentes*. Un conjunto de instancias de propiedades F_1 - F_n realizan una instancia de la propiedad G de un objeto *o sólo si* los objetos en los cuales se instancian F_1 - F_n constituyen o componen x . Esto tiene una implicación muy importante a favor del enfoque dimensionado. Al permitir esta diferencia en objeto y poderes causales entre lo realizado y el realizador, el enfoque dimensionado, según Gillett (2003, 2007), dar cuenta de la relación de realización que se da entre propiedades y los componentes de nivel inferior de los mecanismos que los constituyen. Esto es importante porque este es el tipo de realización más común en las explicaciones mecanicistas de las neurociencias (cf. Craver 2007). Considero que esta es una ventaja crucial del enfoque dimensionado. Una noción adecuada de realización debe poder dar cuenta de estos casos de realización internivel para ser extensionalmente adecuada.

Para clarificar esta noción, podemos ofrecer el ejemplo paradigmático de realización de Gillett (2002): La dureza D de un diamante s^* . s^* es una estructura constituida por los átomos s_1 - s_n , los cuales poseen ciertas propiedades. Poseen, entre otras, las propiedades $L_1, L_2, L_3...$ de estar ligados entre sí y las propiedades $A_1, A_2, A_3...$ de estar alineados de determinada manera. Asumiendo una teoría causal de las propiedades (Shoemaker 2007), las propiedades aportan poderes causales a sus portadores. Recordemos que según el enfoque dimensionado, cada propiedad tiene poderes causales individuativos y lo realizado tiene sus poderes causales en virtud de los poderes causales del realizador. Supóngase, entonces, que entre los poderes causales que la dureza aporta al diamante se encuentra la capacidad de rayar el vidrio. Por su parte, entre los poderes causales que la ligazón y el alineamiento aportan a los átomos de carbono, se encuentra la capacidad de mantener su posición relativa a los otros átomos de

carbono, aun en condiciones extremas. Es claro que el diamante tiene su poder causal en virtud del poder causal mencionado de sus átomos. Por lo tanto, según Gillett (2002), son las propiedades L1, L2, L3... y las relaciones A1, A2, A3..., instanciadas en estos átomos componentes s_1-s_n de s^* , las que conjuntamente realizan la propiedad de la dureza de s^* . Lo importante para diferenciar este enfoque del plano es, entonces, que el realizador y lo realizado son objetos diferentes y que las propiedades causales relevantes del realizado son diferentes de las propiedades causales relevantes del realizador.

Ahora que hemos presentado ambos enfoques quisiera pasar a la evaluación de los mismos. En particular, estoy interesado en dos líneas de crítica diferentes. Las dos son presentadas originalmente como críticas al enfoque dimensionado. Sin embargo, argumentaré que, adecuadamente entendidas, son críticas que obligan a modificar drásticamente el enfoque plano. En la sección 4 mostraré que esta modificación es incluso un elemento fundamental para el que el enfoque plano pueda evitar un problema grave.

3. Mecanicismo y fisicalismo

Gillett mismo propone una primera línea de crítica a la que podría someterse el enfoque dimensionado. Podría objetarse que el realizador de la dureza del diamante no son las instancias de las propiedades de cada uno de sus átomos, sino una instancia de la propiedad estructural, una propiedad que Gillett llama COMBO, de ser una compleja estructura de átomos de carbono ligados y alineados de determinada manera (Gillett 2002). Si admitimos que esta es la propiedad realizadora, este caso ya no difiere de los casos de realización plana en los aspectos que Gillett consideraba relevantes. La propiedad estructural COMBO no sólo está instanciada en el mismo individuo que la dureza (está instanciada en el diamante, como un todo) sino que, muy plausiblemente, un subconjunto de los poderes causales de la primera coinciden con los poderes causales de la segunda. De esta manera, el enfoque plano puede dar cuenta perfectamente del ejemplo de la dureza del diamante sin necesidad de abandonar la idea básica de que todo realizador es una instanciación de la propiedad realizada.

Gillett propone principalmente dos respuestas a esta objeción. En primer lugar, señala que esta estrategia sólo posterga el análisis dimensionado a un nivel inferior. Cada uno de los átomos de carbono que compone la propiedad estructural COMBO está compuesto, a su vez, por otras partículas

fundamentales, como los quarks y los leptones, y estas entidades, a su vez, poseen otras propiedades microfísicas, tales como la carga, el espín, etc. La propiedad COMBO, instanciada en el diamante, no es idéntica a ninguna de estas propiedades microfísicas instanciadas en las partículas fundamentales, aunque tampoco es completamente distinta de ellas. La mejor manera de representar la relación entre COMBO y esos objetos y propiedades microfísicas es, concluye Gillett, en términos de realización dimensionada. Como señaló Endicott (2011), no hay ningún motivo por el que no podamos reiterar un análisis plano mediante propiedades estructurales en el nivel de los objetos y propiedades microfísicas. Podemos decir que la dureza del diamante es realizada por una instancia de una propiedad estructural QUARKO, la propiedad de ser un complejo de quarks, leptones, etc. que poseen tales y cuales propiedades y relaciones fundamentales (Endicott 2011).

Gillett (2007) ha señalado recientemente que el verdadero problema de esta estrategia es que la postulación de propiedades estructurales en cada nivel de un mecanismo implica una violación del criterio científico de parsimonia ontológica. Ni COMBO ni QUARKO son propiedades de los niveles en los que figuran los átomos y las partículas sub-atómicas, respectivamente. Dado un nivel de un mecanismo, esta estrategia parece requerir postular propiedades en ese nivel, adicionales a las que postulan las ciencias que estudian esos niveles. Sin embargo, como ha señalado también Endicott (2011), el hecho de que tales entidades complejas sean “invisibles” desde un determinado nivel científico no implica que no sean “visibles” desde la perspectiva internivel adecuada. Por ejemplo volviendo al caso de Gillett, el enunciado interteórico que determina propiedades de los átomos de un diamante, suficientes para que este tenga su dureza característica, se referirá necesariamente a COMBO. Además, como señala atinadamente Barberis (2014), incluso referentes del mecanicismo, como Craver (2007, p. 212), aceptan la existencia de “propiedades compuestas”, i. e., la propiedad de un mecanismo de tener tales y cuales componentes y actividades organizadas de tal y cual manera. De hecho, resulta muy difícil caracterizar un mecanismo como tal sin apelar a propiedades estructurales.

La segunda objeción que me interesa considerar es una crítica propuesta por Barberis (2014). Barberis propone esta crítica principalmente con el objetivo de mostrar que Fodor y Block, dos de los padrinos de la tesis de la realizabilidad, no adoptan un enfoque dimensionalista, como sostiene Gillett. Sin embargo, considero que esta presentación de la tesis obscurece la fuerza que tiene como crítica al enfoque dimensionado mismo. Barberis sostiene que la (supuesta) adopción del enfoque dimensionado por parte de Block y Fodor es incompatible con el fisicalismo de casos. Barberis presenta dos formulaciones de esta tesis. La primera es la formulación clásica de Fodor según la cual todos los

eventos particulares que caen bajo las leyes de las ciencias especiales (entre ellas, la psicología) son eventos particulares físicos (es decir, eventos que caen bajo leyes de la física) (Fodor 1974). Así, todo caso de un evento o propiedad psicológica es idéntico a un caso de un evento o propiedad física. En segundo lugar, presenta la formulación de Stoljar (2009). Según ella, para todo particular (objeto, evento, o proceso) x efectivamente existente, existe algún particular físico y tal que $x=y$.

Como es sabido, el fisicalismo de casos es una forma de fisicalismo más débil que el fisicalismo de tipos y que, a diferencia de este, es compatible con un dualismo de propiedades. Por ejemplo, puede darse el caso de que para cualquier propiedad psicológica P , P no sea igual a ningún tipo de propiedad física, que no haya ningún tipo de propiedad física que aporte exactamente los mismos poderes causales que P . Puede ocurrir que se de esta situación pero que a la vez todas las instancias de P sean a la vez instancias de propiedades físicas. P puede realizarse en instancias de *diferentes* propiedades físicas A , B , C . Como dichas propiedades son diferentes, no podemos identificar a P con ninguna de ellas. Sin embargo, todos los casos de P son *idénticos a casos* de propiedades físicas. Tal es el requisito mínimo del fisicalismo de casos. Toda propiedad tiene sólo instancias físicas.

Ahora bien, Barberis (2014) considera que en la realización dimensionada puede darse el caso de que la propiedad psicológica P se realice en una instancia de una propiedad física F sin que la instancia de P sea una instancia de F . Pues Barberis considera que hay tres características de la realización dimensionada que hacen al menos difícil mostrar cómo podrían ser instancias idénticas. En primer lugar, estas instancias pueden ser propiedades de diferentes objetos. En segundo lugar, pueden tener poderes causales completamente diferentes. En tercer lugar, los *relata* de la relación de realización son instancias de propiedades. Barberis no es explícito, sin embargo, respecto de por qué estas tres características implican la negación del fisicalismo de casos.

Por mi parte, considero que ninguna de estas características, *por sí solas*, implican la negación del fisicalismo de casos. Creo que si consideramos cada una de estas características, notamos que hacen falta premisas o supuestos adicionales para mostrar dicha incompatibilidad. En primer lugar, meramente suponer que los *relata* de la relación de realización son instancias de una propiedad no implica que no sean la misma instancia de esa propiedad propiedad, que es todo lo que necesita el fisicalismo de casos. En segundo lugar, suponer que dichos *relata* son propiedades de diferentes objetos no implica que no sean la misma propiedad, o incluso la misma instancia de propiedad. Por ejemplo, tener una mancha azul puede ser, en primer lugar, tanto una propiedad de una parte de la pared como de

la totalidad de la pared (es decir, es una misma propiedad de diferentes objetos) y, en segundo lugar, aunque tanto la mancha en la parte de la pared como en la totalidad son instancias de una propiedad (la propiedad de ser una mancha azul) esto no impide que ambas sean la misma instancia de esa propiedad. Por lo tanto, ni la primera ni la tercera condición implican por si mismas un rechazo del fisicalismo de casos.

Considero que respecto de la segunda condición, la posibilidad de que los *relata* de la realización no compartan poderes causales, aunque la situación es más compleja, sucede exactamente lo mismo. Por ejemplo, las instancias de la propiedad de portar la información de que hay un perro son individuadas por cierta relación causal que el objeto que porta dicha información guarda con los perros. Por otro lado, las instancias de la propiedad de ser un dígito, como vimos, no son individuadas por este poder causal (ni por medio de ningún tipo de información). Los dígitos pueden no portar información alguna. La propiedad informacional considerada no comparte poderes causales con la propiedad computacional. Sin embargo, podría darse el caso de que una instancia de la propiedad de ser un dígito sí tenga una relación causal confiable con la propiedad de ser un perro, que el sistema computacional al que pertenece el dígito esté diseñado de forma tal que tal dígito sea causado confiablemente por la presencia de un perro. Parece que aquí podemos decir que esa instancia de la propiedad de ser un dígito es a la vez una instancia de la propiedad de portar información sobre perros. Esto sucede aunque los poderes causales que empleamos para individuar a las instancias de estas propiedades son totalmente diferentes.

Ahora bien, a pesar de que estas tres condiciones no implican por si mismas el rechazo del fisicalismo de casos, considero que la primera y la tercera sí lo hacen si agregamos ciertos supuestos del enfoque dimensionado. Como vimos, el enfoque dimensionado permite que las propiedades realizadoras y realizada sean propiedades de diferentes objetos. En particular, vimos que esto sucede cuando las propiedades realizadoras son propiedades de las diferentes partes componentes del objeto que posee la propiedad realizada. Vimos que en estos casos, las propiedades realizadoras son *varias*, son *diferentes instancias de propiedades* de los diferentes componentes del objeto que posee la propiedad realizada. En el caso de Gillett, los realizadores de la dureza del diamante son diferentes propiedades de cada uno de sus componentes atómicos. Pero entonces la propiedad realizada no puede ser idéntica con las instancias de sus propiedades realizadoras, porque si así fuese estas tendrían que ser idénticas entre sí. La propiedad realizada sólo puede considerarse, tal vez, idéntica con *el conjunto* de propiedades realizadoras, esto es, con COMBO, pero no con cada una de ellas aisladamente. Como

vimos, Gillett niega que la propiedad COMBO constituya el realizador de la dureza del diamante. Por lo tanto, el enfoque de Gillett no permite afirmar que la instancia de la propiedad realizadora (o alguna de las propiedades realizadoras) es idéntica con la propiedad realizada. Considero que este es el verdadero motivo por el que el enfoque dimensionado es incompatible con el fisicalismo de casos.

Lo que más le interesa a Barberis (2013) de esta tensión entre fisicalismo y enfoque dimensionado es que hace implausible que el enfoque dimensionado sea sostenido por uno de los principales defensores del fisicalismo de casos (cf. Fodor 1974, 1981). Por mi parte, considero que implica un problema mucho más grave. Como vimos, el fisicalismo de casos es plausiblemente la única forma de hacer compatible la tesis de la realizabilidad múltiple con el fisicalismo. Como vimos en capítulos anteriores, el fisicalismo, la tesis de que todo lo que existe es, en última instancia, físico, es un supuesto que subyace a toda empresa científica (Fodor 2008), parte del núcleo duro de la actividad científica. Una noción de realización incompatible con tal supuesto es inadmisibles. Esto nos muestra que las dos críticas consideradas están conectadas. Por un lado, *no tenemos buenas razones* para considerar que en los casos de realización presuntamente dimensionada, el realizador no sea una propiedad como COMBO, y que por lo tanto esos sean en realidad casos de realización plana. En segundo lugar, *tenemos buenas razones* para aceptar que en estos mismos casos el realizador es efectivamente una propiedad de tipo COMBO. Si no lo hacemos, la realización es incompatible con el fisicalismo de casos.

Ahora bien, según el criterio ofrecido para la realización plana, se sigue que la realización puede darse entre propiedades como COMBO y QUARKO. Después de todo COMBO y QUARKO son propiedades del mismo individuo y comparten poderes causales. Pero esto implica que este tipo de realización se puede dar entre propiedades que parecen pertenecer a diferentes niveles (en este caso la relación se da entre un mecanismo cuyos componentes son átomos y un mecanismo cuyos componentes son las partículas sub-atómicas que componen a dichos átomos). Si esto es así, el enfoque que adoptamos ya no sería estrictamente plano. La realización de COMBO en QUARKO es un caso de realización plana dimensionada. Es plana en que se da en un mismo objeto y en que los *relata* comparten poderes causales y es dimensionada en que relaciona elementos que parecen pertenecer a diferentes niveles. De este modo, las dos críticas que apuntan a favor de la adopción de este tipo de propiedades como *relata* de la realización, no sólo conllevan el abandono de supuestos del enfoque dimensionado sino también de supuestos relevantes para el enfoque plano.

Ahora bien, lamentablemente esta conclusión es aún prematura. Pues hay buenas razones para pensar que la noción de nivel presupuesta en lo recién dicho no es la más adecuada y que la noción que debemos aceptar no nos permite sacar las mismas conclusiones. Barberis (2014) ha discutido un problema adicional, un tercer problema, para el enfoque dimensionado. Considera que la aceptación de la transitividad de la realización por parte del enfoque dimensionado genera casos espurios de múltiple realización. Previendo esta línea de crítica, Gillett ha propuesto una especificación de su enfoque en términos de la noción de nivel. Barberis (2014) cuestiona la caracterización de Gillett de la noción de nivel y sugiere que la noción adecuada es la noción mecanicista propuesta por Craver (2007). Barberis argumenta que esta noción de nivel permite que el enfoque plano pero no el dimensionado evite los casos de realización múltiple espuria. En la sección siguiente presentaremos brevemente esta discusión. En segundo lugar, argumentaré que la noción adecuada de nivel, en contra de lo que argumenta Barberis, permite que el enfoque dimensionado evite la realización múltiple espuria pero impide que el enfoque plano lo haga. En tercer lugar, mostraré que el enfoque plano puede evitar estos casos sólo si reparamos que, en cierto sentido, relaciona propiedades que pertenecen a diferentes niveles, según la noción de Craver (2007), y que por lo tanto es y debe ser, en cierto sentido, dimensionado.

4. Múltiple realizabilidad espuria y la dimensionalidad de lo plano

Barberis (2014) argumenta que el enfoque dimensionado enfrenta un problema grave. Implica casos espurios de realización múltiple, casos que no constituyen aplicaciones pretendidas de la noción de realización. La tesis de Barberis es que este problema sólo puede evitarse si se abandona la transitividad de la relación de realización.

De acuerdo con los defensores de la visión dimensionada, la realización es una relación transitiva. Aizawa y Gillett (2009a, p. 194) afirman que “una divergencia adicional [con el enfoque plano] es que según nuestro enfoque la realización científica es una relación transitiva, mientras que bajo la propuesta de Shapiro la realización no es transitiva.” La transitividad implica que, si B realiza A, y C realiza B, entonces C realiza A. Como señala Barberis, se supone que esta tesis articula la intuición de que la realización es un tipo de determinación. Así, si B es metafísicamente suficiente para A, y C es metafísicamente suficiente para B, entonces es esperable que C sea suficiente para A también. Retomando el ejemplo de Gillett, supongamos que cierta propiedad *P* de alguna de las partículas

microfísicas que componen realizan alguna propiedad Q de algún átomo de carbono, y supongamos que a su vez, que Q realiza la propiedad de la dureza D de un diamante. Dada la transitividad de la realización, se sigue que la propiedad P realiza la propiedad D del diamante (Aizawa y Gillett 2009a).

Uno de los problemas que inmediatamente implica este enfoque y que Aizawa y Gillett mismo notaron, es que permite la aparición de casos espurios de múltiple realización (Aizawa y Gillett 2009a, b). En el caso anterior D es realizado tanto por P como por Q . Es decir, D se realiza múltiplemente. Esta es claramente una aplicación no pretendida de la idea de múltiple realización. Como señala Barberis, las noción de múltiple realización supone que los realizadores están, en algún sentido, en el mismo nivel. Si este ejemplo se aceptase como un caso genuino de múltiple realización, entonces todas las propiedades reales que no estén realizadas directamente en el nivel fundamental estarían, *ipso facto*, múltiplemente realizadas.

Esta idea de recurrir a la noción de nivel surge inmediatamente al considerar el modo de restringir la noción de realización de manera tal que no permita estos casos. Gillett y Aizawa restringen de este modo la noción afirmando que una propiedad G está múltiplemente realizada si y sólo si: (i) bajo las condiciones $\$$, un individuo s posee una instancia de una propiedad G en virtud de los poderes causales aportados por las instancias de las propiedades o relaciones F_1 - F_n a s , o a los constituyentes de s , pero no viceversa; (ii) bajo las condiciones $\* un individuo s^* posee una instancia de una propiedad G en virtud de los poderes causales aportados por las instancias de las propiedades o relaciones F^*_1 - F^*_m a s^* , o a los constituyentes de s^* , pero no viceversa; (iii) las propiedades o relaciones F_1 - F_m son distintas de las propiedades o relaciones F^*_1 - F^*_n ; (iv) bajo las condiciones $\$$ y $\* , las propiedades F_1 - F_n y F^*_1 - F^*_m pertenecen al mismo nivel científico (Aizawa y Gillett 2009a, p. 188, la nomenclatura es de Barberis).

La condición relevante para nuestro problema es, obviamente, la última. Esta condición exige de manera explícita que los realizadores, para ser *múltiples*, deben pertenecer al mismo nivel científico. Según Aizawa y Gillett, se trata de un presupuesto común a todos los análisis de la múltiple realización, aunque raramente explicitado. Sin embargo, Barberis cree que el enfoque plano no explicita esta condición porque no la necesita. Barberis afirma que como la propiedad realizada y su realizador están instanciados en el mismo objeto y comparten poderes causales, se sigue inmediatamente que todos los realizadores potenciales pertenecen al mismo nivel. La necesidad de la condición (iv) es idiosincrática del enfoque dimensionado. Ahora bien, considero que la discusión que a continuación desarrolla

Barberis nos muestra justamente que el hecho de que el enfoque plano no suponga niveles no se sigue “inmediatamente”, como señala Barberis, sino que se sigue de la noción específica de nivel que el autor va a defender.

Una primera pregunta que podemos hacernos respecto de la condición de Aizawa y Gillett es qué noción de nivel presupone. Según ellos, dos entidades pertenecen al mismo nivel científico cuando participan, *o pueden participar* en el mismo mecanismo, bajo ciertas condiciones, o bien cuando participan en procesos que conjuntamente implementan otros procesos (2009a, p.43, n. 11). Barberis (2013) argumentará que este análisis de los niveles en ciencia presupone el enfoque de Wimsatt (1994) sobre el mismo tema. Sin embargo, argumenta Barberis, el empleo de la noción de “mecanismo” por parte de Aizawa y Gillett parece presuponer también una noción mecanicista de nivel (Craver 2007, Bechtel y Richardson 2010; cf. Barberis 2012), la noción de nivel empleada en el contexto de la explicación mecanicista. Ahora bien esta noción difiere notablemente de la de Wimsatt (1994).

Wimsatt representa a los niveles como máximos locales de interacción causal, de regularidad y de predictibilidad en diferentes escalas de tamaño. Este autor sostiene una hipótesis empírica según la cual los picos de interacción y regularidad se producen entre items que pertenecen a un mismo nivel y no entre items de diferentes niveles. Los objetos de un nivel interactúan con mayor fuerza y más frecuentemente con los otros objetos que pertenecen al mismo nivel. Barberis señala muy agudamente que la misma intuición parece estar presupuesta en el análisis de Aizawa y Gillett cuando afirman que dos objetos pertenecen al mismo nivel cuando *pueden* implementar conjuntamente un determinado proceso causal. Existe algo acerca de los objetos de un mismo nivel que les posibilita interactuar entre ellos de manera tal de implementar procesos causales. Según Wimsatt, la característica relevante es que se trata de objetos de, más o menos, el mismo tamaño, sobre los que pueden actuar, en consecuencia, las mismas fuerzas.

Barberis cree que hay buenas razones para pensar que el enfoque de Wimsatt no constituye un enfoque adecuado acerca de los niveles en el contexto de la explicación mecanicista. La noción de nivel que está involucrada en los niveles de mecanismos parece tener un alcance mucho más local, es decir, no presupone que hay grandes divisiones entre estratos compuestos por tipos de propiedades. Craver caracteriza a los niveles de la siguiente manera:

En los niveles de mecanismo, los *relata* son mecanismos actuantes en niveles superiores y sus

componentes en niveles inferiores. Estos *relata* no se conciben propiamente como entidades ni como actividades; más bien, deberían entenderse como entidades actuantes. La relación internivel es la siguiente: el hacer- ϕ por parte de X está en un nivel de mecanismo inferior al hacer- ψ por parte de S si y sólo si el hacer- ϕ por parte de X es un componente en el mecanismo para hacer- ψ por parte de S. Los componentes de nivel inferior se *organizan conjuntamente* para formar los componentes de nivel superior. (Craver 2007, p. 189)

Se ve entonces en qué sentido la noción de nivel tiene un alcance más restringido, más local. Esta caracterización implica que los niveles de mecanismos sólo están definidos en el contexto de una determinada jerarquía composicional. Un objeto (ejecutando una determinada actividad) A está en un nivel de mecanismo inferior a otro objeto (ejecutando otra actividad) B si y sólo si *A es un componente en el mecanismo que subyace a B*. Esto quiere decir que la jerarquía de niveles puede ser distinta para diferentes mecanismos. Cuántos niveles haya, y qué niveles se incluyan en la jerarquía, dependerá de qué componentes sean explicativamente relevantes en cada caso. Los niveles sólo tienen relevancia local (Craver 2007, p. 191). No tenemos nada como un nivel biológico, químico, atómico, etc. de los que podamos decir si son superiores o inferiores a otro nivel dado, en general, independientemente de un mecanismo particular.

De esto se sigue que la noción mecanicista es incompatible con la que parecen presuponer Aizawa y Gillett en su análisis de la múltiple realización. A continuación, Barberis (2014) argumenta que la noción mecanicista no le permite al enfoque dimensionado excluir los casos espureos de realización, pero sí al enfoque plano. Considero que este diagnóstico es errado, que lo que sucede es exactamente lo contrario: la noción de Craver le permite al enfoque dimensionado, pero no al plano, excluir los casos espureos. Veamos brevemente la argumentación de Barberis.

Barberis hace considerar el siguiente caso: Supongamos que tenemos un típico explanandum de la explicación mecanicista: la capacidad psicológica D, la capacidad de ese individuo de consolidar una experiencia en la memoria de largo plazo. Supongamos que se descubre que a D subyace el mecanismo neurobiológico B1. Como muchos mecanismos neurobiológicos, B1 está constituido por un denso arreglo de diversas poblaciones neuronales, vías metabólicas, distintos tipos de macromoléculas, canales iónicos, iones específicos, genes, etc. organizados de determinada manera. Pero supongamos además que se descubre otro mecanismo neurobiológico B2 como posible realizador de la capacidad D.

B2 es también un arreglo complejo y abigarrado de poblaciones neuronales, genes, vías metabólicas, etc., aunque distinto del mecanismo B1. Como vimos, según Aizawa y Gillett, para determinar que trata de un caso genuino de múltiple realización, y no de múltiples realizaciones espurias debemos determinar que B1 y B2 pertenecen al mismo nivel. Pero, argumenta Barberis, dado que B1 y B2 son mecanismos distintos, no es posible determinar, según la concepción mecanicista de nivel, si los objetos y las actividades que los componen pertenecen, o no, al mismo nivel de mecanismos.

Aún cuando B1 y B2 estén compuestos de los mismos tipos de objetos, aún cuando entre los componentes de B1 y de B2 se incluyan, por ejemplo, canales de potasio, de eso no se sigue que dichos canales pertenezcan al mismo nivel en uno y otro mecanismo. No hay nada intrínseco acerca de esos objetos que nos permitan determinar si pertenecen al mismo nivel en uno y otro mecanismo. Pues la determinación de niveles, según este enfoque, es relativa a cierta jerarquía composicional para un mecanismo dado: Si son partes de un mismo mecanismo tales que una de ellas es un componente de la otra, entonces están en diferentes niveles, si son partes del mismo mecanismo tales que ninguna es componente de la otra, entonces pertenecen al mismo nivel. Por el contrario, los componentes de B1 y B2 ni siquiera son componentes del mismo mecanismo. Por lo tanto, ni siquiera tiene sentido la pregunta de si pertenecen a diferentes niveles o no.

Por el contrario, argumenta Barberis, en la concepción plana de la realización, está garantizado que B1 y B2 son del mismo nivel. La idea de Barberis es que como los realizadores, según el enfoque plano, son las propiedades estructurales de ser un mecanismo B1 y ser un mecanismo B2 y como estas propiedades estructurales están en el mismo nivel que la propiedad realizada D (pues ni B1 ni B2 son componentes de D , sino que D y B1, por un lado, y D y B2, por el otro, son simplemente diferentes propiedades de un mismo mecanismo u objeto) se sigue entonces que B1 y B2 están en el mismo nivel.

Ahora bien, considero que estos dos argumentos no funcionan y que la situación es la inversa a la que presenta Barberis: el criterio de Craver permite excluir a la realización espuria desde el enfoque dimensionado pero no desde el plano. Empecemos considerando el plano. Pensemos en un caso no como el de B1 y B2 sino un caso de realización múltiple espuria. Retomemos el ejemplo de las propiedades COMBO y QUARKO que realizan la dureza D de un diamante. Considero que el enfoque plano (aunque, como vimos, no el dimensionado) puede y debe admitir tanto COMBO como QUARKO como a realizadores de la propiedad D . Podemos pensar que si aceptamos, con Barberis, que la realización plana no es transitiva, entonces el enfoque plano no debe admitir a QUARKO como

realizador de *D*. Ahora bien, considero que (i) el modo en que enfoque plano es incompatible específicamente con transitividad *permite* que tanto QUARKO como COMBO sean realizadores de *D*, y que (ii) el enfoque plano *debe* afirmar que tanto COMBO como QUARKO realizan a *D*.

Empecemos por (i). Vimos que, respecto de la transitividad, lo que el enfoque plano no permite es que QUARKO realice a COMBO si COMBO realiza a *D*. Pues dado que *COMBO*, al ser realizador de *D*, es un *particular o instancia*, no puede ser una propiedad realizada por QUARKO. Como vimos, según el enfoque plano, lo realizado es una propiedad funcional, abstracta, no un particular o una instancia. Pero esto no implica que según el enfoque plano la transitividad de la realización sea falsa, sino sólo que nunca se cumple su antecedente. La transitividad afirma que si A es realizado por B y B por C entonces A es realizado por C. Lo que impide el enfoque plano es que si A es realizado por B, B sea realizado por C. Impide que se de el antecedente del principio de transitividad, que se de el segundo conyunto, si se da el primero. Pero esto es compatible con que B y C realicen a A, siempre que C no realice a B (o viceversa). En particular, es compatible con que QUARKO y COMBO realicen a *D*, siempre que QUARKO no realice a COMBO, o viceversa.

Ahora bien, considero que aún aceptando la no transitividad, el enfoque plano *debe* afirmar que tanto COMBO como QUARKO son realizadores de *D*. Pues el criterio que determina a algo como realizador de un rol funcional *D* es, como vimos, que sea una propiedad que tenga a dicho rol funcional como un subconjunto de sus propiedades causales. QUARKO claramente cumple con este requisito, pues no es otra cosa que la propiedad estructural de ser el conjunto de todas las partículas sub-atómicas y propiedades suyas que componen al diamante. Por lo tanto, el enfoque plano no es sólo compatible con que QUARKO y COMBO sean realizadores de *D*, sino que debe afirmar que este es el caso.

Ahora bien QUARKO y COMBO, además de ser realizadores diferentes de *D*, son semejantes a B1 y B2 del ejemplo anterior en que ambos están en el mismo nivel que lo realizado *D*. Al igual que B1 y B2, QUARKO y COMBO son propiedades del mismo objeto del que *D* es una propiedad, no son componentes de *D*. Por lo tanto, siguiendo a Barberis, tenemos que decir que están en el mismo nivel y que constituyen un caso genuino de realización múltiple. Esto quiere decir que, en contra de Barberis, el criterio de Craver no permite que el enfoque plano excluya los casos de realización múltiple espuria. Pasemos ahora a ver qué sucede con el enfoque dimensionado.

Barberis señaló adecuadamente, a mi parecer, que como los diferentes realizadores la capacidad psicológica *D*, los componentes de B1 y B2, respectivamente, son componentes de mecanismos

diferentes y además pertenecen (según el criterio mecanicista) a un nivel diferente al de D (pues tanto las partes de $B1$ como las de $B2$ son componentes de D), no podemos determinar si $B1$ o $B2$ están en el mismo nivel. Más aún, dicha pregunta ni siquiera tiene sentido en este marco. Ahora bien, considero que no necesitamos responder a esa pregunta para excluir los casos de realización espuria, no necesitamos poder determinar que dos realizadores pertenecen al mismo nivel para excluir a los casos problemáticos. Pues nos basta con que podamos decir cuándo dos realizadores *no* pertenecen al mismo nivel. Creo que esto es algo que el enfoque dimensionado sí puede hacer. En el caso de QUARKO y COMBO, los realizadores de la dureza del diamante no serán, obviamente QUARKO y COMBO mismos, sino las propiedades de sus partes componentes. Ahora bien, es claro que los componentes de QUARKO pertenecen a un nivel inferior respecto de los de COMBO, pues los componentes de QUARKO no son otra cosa que las partículas sub-atómicas (y sus propiedades) que componen a los átomos que constituyen COMBO y que componen a su vez a la dureza D . El enfoque dimensionado puede excluir a este caso como uno de realización múltiple por el hecho de que los dos tipos de realizadores pertenecen a diferentes niveles. La condición de Aizawa y Gillett no necesita decir que los realizadores de los casos genuinos pertenecen al mismo nivel, sino, negativamente, *que no pertenecen a niveles distintos*. Así, adoptando la noción mecanicista de nivel, el enfoque dimensionado puede excluir los casos de realización múltiple espuria.

De este modo, inversamente a lo que sostiene Barberis (2013) la noción mecanicista de nivel le permite al enfoque dimensionado pero no al plano excluir los casos de realización múltiple espuria. Ahora bien, esto implica un problema. A partir de esta conclusión no podemos simplemente adoptar el enfoque dimensionado ya que, como vimos en la sección anterior, este enfoque es incompatible con el fisicalismo de casos. De este modo, nos encontramos con un dilema: Si adoptamos el enfoque dimensionado perdemos el fisicalismo de casos, pero si adoptamos el enfoque plano, no podemos excluir los casos de realización múltiple espuria. Creo, sin embargo, que podemos aún refinar de un modo natural la condición que define la realizabilidad múltiple de modo tal que el enfoque plano pueda también excluir la realizabilidad múltiple espuria.

Vimos que según la caracterización de Craver (2007) los niveles están determinados por una jerarquía composicional. Los relata de la composicionalidad son entidades actuantes y el hacer P por parte de r es de nivel inferior a hacer Q por parte de s sólo si el primero es un componente del segundo. Los componentes de nivel inferior se organizan conjuntamente, según Craver, para formar los componentes de nivel superior. Como vimos, los realizadores en un caso de realización múltiple

espuria para el enfoque plano son propiedades estructurales como QUARKO y COMBO. No podemos afirmar, según el enfoque de Craver, que estas propiedades están en diferentes niveles porque ninguna es un componente de la otra. QUARKO es el conjunto de partículas sub-atómicas (y sus propiedades y relaciones) que conjuntamente forman un diamante y COMBO es un conjunto de partículas atómicas (y sus propiedades y relaciones) que conjuntamente forman o constituyen el mismo diamante.

Ahora bien, hay sin embargo algún tipo de relación composicional entre COMBO y QUARKO. Después de todo, los componentes de QUARKO son componentes que conjuntamente forman los componentes de COMBO. Los componentes de QUARKO son las partículas sub-atómicas (más propiedades y relaciones) que forman los átomos que componen COMBO. Si bien QUARKO y COMBO no pertenecen a diferentes niveles, según la noción mecanicista, sus componentes sí lo hacen. De este modo, podemos hacer que el enfoque plano excluya a la realizabilidad múltiple espuria afirmando que hay realizabilidad múltiple sólo si no podemos decir que *los componentes* de los diferentes realizadores (y no meramente los realizadores mismos) pertenecen a diferentes niveles de mecanismo.

Una consecuencia importante de considerar al caso COMBO/QUARKO de este modo es que, si bien ahora el enfoque plano puede excluir a este caso como un caso de realización múltiple, la realización plana ya no es genuinamente plana. Si bien la propiedad realizada, la dureza D del diamante, y las propiedades realizadoras, QUARKO y COMBO, no pertenecen a diferentes niveles, como vimos, sus componentes sí lo hacen *y este es un factor constitutivo de la realizabilidad*. Por lo tanto, la realización plana involucra necesariamente, elementos que pertenecen a diferentes niveles. Al final de la sección anterior consideré que el enfoque plano se volvía dimensionado en cierto sentido al aceptar propiedades como QUARKO y COMBO entre los *relata* de la realización. Recién ahora, habiendo discutido la noción de nivel y habiendo argumentado la importancia de integrarla al enfoque plano, podemos afirmar que efectivamente la noción de realizabilidad, si bien conservando supuestos centrales del enfoque plano, incorpora un elemento típicamente dimensionado.

En la próxima sección consideraré qué relaciones podemos determinar, según esta noción, entre el procesamiento neuronal, la computación y la cognición. Además, argumentaré reflexionar sobre el caso de la relación entre computación y cognición, si bien no nos alejan de la definición plana, si nos muestra que otro de los presuntos supuestos de esta definición es también prescindible. En particular, no necesitamos que los *relata* compartan poderes causales.

5. Cognición, computación y procesamiento neuronal: realización como determinación nomológica o metafísica

Hemos visto en las secciones anteriores que el enfoque más plausible sobre la realización es una mezcla de realización dimensionada y plana. De la propuesta plana tomamos las tesis de que la propiedad realizadora es una instancia de la propiedad realizada. Aceptar esto, como vimos, no es opcional, ya que de ello depende que la noción de realización sea compatible con el fisicalismo de casos. De la realización dimensionada tomamos la tesis de que la relación de realización puede relacionar elementos que pertenecen a diferentes niveles. De esto depende que podamos excluir a los casos espureos de realizabilidad múltiple.

Ahora bien, aclarado un poco más el panorama respecto de lo que constituye la relación de realización, estamos tal vez en una mejor posición para preguntarnos sobre la relación metafísica que guardan la cognición, la computación y el procesamiento neuronal. Empecemos por los primeros dos.

He argumentado en los capítulos 7 y 8 que la noción de procesamiento de información no implica conceptualmente la noción de computación, ni viceversa. Vimos en el capítulo 8 que Fodor (1994) había defendido esta tesis y había argumentado que tiene consecuencias para la relación de realización entre la cognición y la computación. Por mi parte sostuve que el argumento de Fodor no puede funcionar con una noción de computación como la que posee. La noción de forma sintáctica no permite caracterizar adecuadamente a la computación. Por el contrario la noción de independencia del medio de Piccinini y Bahar sí parece ser más adecuada, ya que rescata la característica común a los dos tipos más generales de computación (digital y analógica) y se trata de una característica que no posee cualquier sistema físico. Este enfoque nos permite recoger todos los sentidos computación relevantes y nos permite distinguir entre sistemas físicos computacionales y no computacionales. Sin embargo, Piccinini y Bahar sostiene que su noción de computación sí es implicada por la noción de información. Por el contrario, he mostrado que esta implicación no se da. Brevemente, argumenté que la independencia del medio que poseen los estados informacionales es tal que, un tipo de estado informacional puede instanciarse en un objeto físico cuyas propiedades intrínsecas no sean independientes del medio, en el sentido computacional. Como las propiedades que definen a los estados informacionales son puramente extrínsecas o relacionales, no hay ninguna restricción para las

propiedades intrínsecas que puedan tener sus instancias. Por su parte, para que un vehículo sea computacional, sus propiedades intrínsecas causalmente relevantes tienen que ser abstractas, tienen que ser propiedades físicas no muy específicas. De este modo, la instanciación de un estado informacional no implica la instanciación de un vehículo computacional. Como un estado informacional puede instanciarse en un estado físico de cualquier tipo, puede instanciarse, en particular, en uno cuyas propiedades funcionalmente relevantes no sean independientes del medio. Que la implicación en la otra dirección no se da es algo que ni Piccinini y Bahar ponen en duda. Un sistema de computación digital, por ejemplo, puede manipular dígitos que no porten información alguna sobre estados externos al sistema.

Dado el enfoque que hemos adoptado sobre la realización, esta falta de relación conceptual parece decirnos algo sobre la relación metafísica entre computación e información. En la sección anterior argumenté que el supuesto dimensionado de que los *relata* de la realización pueden no compartir poderes causales no es incompatible con el fisicalismo de casos, no es problemático en este sentido. Sin embargo, tal vez sí sea un supuesto incompatible con el enfoque plano, tal vez sea un supuesto exclusivamente dimensionado. Esto es por lo siguiente. Como hemos visto, la definición plana de realización es la relación entre una propiedad funcional P y otra propiedad Q que *ocupa el rol funcional* P . Al menos un modo de entender la idea de ocupar un rol funcional (Shoemaker 2007) es afirmar que Q ocupa el rol funcional P ssi los poderes causales que constituyen P son un subconjunto de los poderes causales que constituyen Q . De esto se sigue que, según esta concepción, si los poderes causales de Q no incluyen a los poderes causales de P entonces Q no ocupa dicho rol funcional y por lo tanto no realiza dicha propiedad.

¿Es necesario entender a la realización de este modo, como la relación de dos propiedades que comparten poderes causales? En la sección anterior argumenté que casos de propiedades que no comparten poderes causales, no son incompatibles con el fisicalismo de casos. Podemos preguntarnos ahora si son incompatibles con la noción de realización. El argumento de la sección pasada presentaba, justamente el caso de una instancia de vehículo computacional y una instancia de estado informacional ¿Podemos decir que, a pesar de que estas propiedades no comparten poderes causales, se realizan la una a la otra, que, por ejemplo, el dígito realiza al estado informacional?

Considero que el problema principal para considerar en casos como este, donde no hay solapamiento de poderes causales, como casos de realización es que resulta difícil determinar cuál es el

realizador y cuál el realizado. En los casos planos típicos, el hecho de que lo realizado sea un *subconjunto* de los poderes causales del realizador genera una asimetría que nos permite distinguir claramente al realizador del realizado. En este caso tenemos dos propiedades que se instancian juntas o que constituyen la misma instancia, pero no tenemos la asimetría propia de la relación de realización. La propiedad de ser un dígito no contiene como parte propia nada de lo que *constituye* la de ser un estado informacional ni viceversa. No tenemos un criterio claro para decidir si es el caso de que la instancia de dígito realiza la propiedad de ser un estado informacional, o si la instancia de estado informacional realiza la propiedad de ser un dígito.

Creo que la asimetría de la realización puede obtenerse para estos casos si tomamos la idea de Fodor (1995) de que la relación de determinación entre realizador y realizado no necesita ser metafísica sino que puede ser nomológica. No necesitamos que en todo mundo posible en que se instancia el realizador, se instancie el realizado (pero no viceversa). Nos basta con que en todo mundo nomológicamente posible (en que se den ciertas condiciones *C*) en que se instancie el realizador, se instancie el realizado (pero no viceversa). Tomando esta idea, podemos ver que hay un sentido en que los vehículos computacionales y los informacionales pueden compartir poderes causales y por lo tanto, un sentido en que los primeros pueden ocupar los roles funcionales que constituyen los segundos. Podemos decir que una propiedad *P* ocupa el rol funcional *Q* si los poderes causales de *Q* son un subconjunto de los poderes causales de *P* en los mundos nomológicamente posibles (y en las condiciones *C*). En dichas condiciones, los vehículos computacionales no tienen, como parte de sus propiedades causales, sólo a sus propiedades causales constitutivas, sino también a las que constituyen a un estado informacional. Así, podemos aplicar la noción plana de realización al caso de la computación y la información o cognición si permitimos esta restricción para la noción de ocupar un rol funcional. De este modo, si bien, de acuerdo con Shoemaker, compartir poderes causales en general, es necesario para que haya una relación de realización entre dos propiedades, no es necesario que estos poderes sean *constitutivos* de ambas propiedades. La noción plana de realización es compatible con que los *relata* no compartan poderes causales en este sentido más fuerte. Basta con que el realizador tenga los poderes causales de lo realizado *sólo en ciertas condiciones, en los mundos nomológicamente posibles*.

Ahora podemos pasar a considerar la relación entre procesamiento neuronal y computación y procesamiento neuronal y cognición. Como hemos visto, a diferencia de las nociones de computación y cognición, estos dos pares de nociones sí guardan relaciones conceptuales. El procesamiento neuronal

implica tanto computación como cognición. Hemos visto que esta implicación se da en cada caso por características diferentes del procesamiento neuronal. La computación es implicada por la independencia del medio de las tasas y tiempos de los potenciales de acción neuronal. La cognición es implicada por el tipo complejo de retroalimentación negativa que lleva a cabo el sistema nervioso humano. Vimos también que estas dos relaciones conceptuales apuntaban, aparentemente, a relaciones metafísicas diferentes. Por un lado, vimos que ser un vehículo del procesamiento neuronal es simplemente ser un tipo específico de computación, no un realizador de la misma. La idea detrás de esta intuición es que la noción de computación se aplica a la de procesamiento neuronal en el mismo sentido en que se aplica a las nociones de computación digital y analógica. Son simplemente tipos de procesamiento que tienen las propiedades generales del procesamiento computacional. Si decimos que el procesamiento neuronal realiza al procesamiento computacional también debemos decir lo mismo de la computación digital y analógica.

Esta tesis estaba basada aún en una comprensión pre-teórica de la noción de realización. Ahora que hemos considerado ampliamente dicha noción, podemos dar una justificación más fundada para la tesis de que no hay una relación de realización entre computación y procesamiento neuronal. La razón es principalmente que la computación genérica no es caracterizada en términos funcionales. Vimos que para generar la relación adecuada entre realizador y realizado, al menos lo realizado debe caracterizarse funcionalmente, por medio de sus poderes causales. Sin embargo, la noción de vehículo de la computación genérica es caracterizada meramente en base a las propiedades intrínsecas que son relevantes para sus poderes causales, *y no especificando algo sobre dichos poderes*. Sólo se dice que las propiedades causalmente relevantes de los vehículos informacionales son abstractas, pero no se determina qué poderes causales específicos tienen dichas propiedades. Adoptando el enfoque plano de la realización entonces, podemos justificar la tesis de que la computación genérica no es realizada por la computación neuronal, sino que esta es meramente un tipo de aquella.

La relación entre procesamiento neuronal e información o cognición es, como vimos diferente. Aunque aquí también tenemos una relación conceptual, los relata son ambos caracterizados en términos puramente funcionales o causales. Lo realizado, la información y la cognición es caracterizada en términos de relaciones causales de estructuras neuronales con estados externos y relaciones causales de esas estructuras entre sí (recordemos que el procesamiento de información requiere que el sistema *responda a las diferentes estructuras* según la información que portan). Por su parte, las diferentes nociones de retroalimentación neuronal que consideramos involucran diferentes formas de loops

causales entre valores de variables. Más aún, la relación que guardan la información y la retroalimentación neuronal es la forma más fuerte en que se relaciona un rol funcional y su ocupante. Los poderes causales *constitutivos* del procesamiento de información son parte de los poderes causales *constitutivos* de la retroalimentación nerviosa.

Así, dadas la reflexión sobre la noción de realización obtenemos tres relaciones diferentes entre procesamiento de información, computación y procesamiento neuronal. El procesamiento de información se realiza en la computación de un modo metafísicamente contingente, la computación no se realiza en el procesamiento neuronal sino que simplemente es un tipo específico de computación y el procesamiento neuronal realiza al procesamiento de información de un modo metafísicamente necesario (es decir, es metafísicamente suficiente para la instanciación de un proceso informacional). Considero que estas tres relaciones nos dejan una lección interesante respecto de la relación entre la realización de la cognición y el fisicalismo respecto de la misma, que era el objetivo de esta segunda parte.

Vimos que Barberis había argumentado que la transitividad de la noción de realización nos impedía excluir casos de realización múltiple espuria. Por mi parte he argumentado, que la transitividad no era responsable de esos casos (pues pueden generarse para un enfoque que, como el plano, no implique transitividad) sino que necesitábamos, como creía Gillett, una noción de nivel. Ahora bien, podemos pensar que la ausencia de transitividad no sólo no nos evita un problema, respecto de la realizabilidad espuria, sino que nos trae uno más grave, la imposibilidad de dar cuenta del fisicalismo. Pues supongamos que tenemos una propiedad psicológica. De ella queremos decir que se realiza en un estado computacional, pero como somos fisicalistas, también queremos decir que se realiza en un estado físico. Parece que podemos decir esto último, justamente *porque el estado computacional se realiza en un estado físico*. Por lo tanto, los estados psicológicos serían físicos sólo si vale transitividad, si el hecho de que un estado psicológico se realice en uno computacional y éste en uno físico, implica que el estado psicológico también se realiza en un estado físico.

El enfoque propuesto nos muestra que no hace falta transitividad para adoptar una perspectiva fisicalista respecto de la realización de los estados psicológicos. Pues si caracterizamos a lo realizado en términos funcionales, lo único que necesitamos para que se realice en diferentes niveles es que haya propiedades en esos niveles que ocupen ese rol funcional. Lo cognitivo puede realizarse *directamente* sin la mediación de otra relación de realización, en diferentes niveles. En particular, puede realizarse en

lo computacional y en lo neuronal sin que lo computacional mismo se realice en lo neuronal. Así, la no transitividad de la realización no es, a pesar de las apariencias, un problema para el fisicalismo en general y el fisicalismo respecto de los estados psicológicos en particular. Ahora bien, este tipo de realización física sin transitividad no es sólo una posibilidad lógica posibilitada por nuestra noción de realización sino que es, como vimos, lo que efectivamente ocurre. Los procesos cognitivos efectivamente se realizan directamente sobre una base computacional y neuronal, aunque, como argumenté, lo computacional no se realiza sobre lo neuronal.

6. Conclusión

En este capítulo me he interiorizado en la discusión contemporánea en torno a la noción de realizabilidad para, argumentado a favor de una perspectiva respecto de varios aspectos de dicha noción, determinar si se aplica y de qué modo a la relación entre las propiedades sobre las que reflexionamos en los capítulos anteriores. En primer lugar, he ofrecido lo que considero una versión más clara del argumento de Barberis (2014) de que el enfoque dimensionado es incompatible con el fisicalismo de casos. Superar este problema supone aceptar propiedades estructurales como COMBO o QUARKO como *relata* de la realización. En segundo lugar he mostrado que para que el enfoque plano pueda excluir los casos de realización múltiple espuria debemos considerar que en muchos de los casos de realización que involucran propiedades de tipo COMBO la realización vincula propiedades que pertenecen a diferentes niveles. Esto hace que el enfoque no pueda ser llamado genuinamente “plano”.

Por último, mostré que podemos aplicar la noción de ocupar un rol funcional a propiedades que no comparten poderes causales *constitutivos* y que por lo tanto, dichas propiedades pueden encontrarse en una relación de realización, según el enfoque plano. Este es el caso de la relación entre cognición y computación. Cuando esto sucede, el realizador no es metafísica sino nomológicamente suficiente para que se de lo realizado. Por lo tanto, las tres nociones consideradas en los capítulos anteriores guardan tres relaciones diferentes entre sí. La cognición se realiza sobre una base computacional de un modo metafísicamente contingente, la computación es simplemente un género del que el procesamiento neuronal es un tipo específico y el procesamiento neuronal realiza a la cognición de un modo metafísicamente necesario.

Estas tres tesis constituyen el mapa de relaciones por el que podemos finalmente determinar de un modo preciso cómo se relaciona la cognición, entendida informacionalmente, con su base física específica (computacional y neuronal).

Capítulo 11: Conclusión. El sistema de la mente informativa

He desarrollado en las secciones anteriores una serie de elucidaciones guiadas por medio de dos ideas: La idea de que la cognición es procesamiento de información y la de que la cognición así entendida permite explicar de qué modo es realizada en los niveles inferiores computacional y neuronal. Es momento de recapitular los puntos centrales de esta tarea y presentar la trama unificada de relaciones que la información permite establecer entre los diferentes aspectos que constituyen la mente.

En la primera parte me he ocupado de la elucidación de las nociones de concepto, representación perceptual y conocimiento y su relación con la noción de información. En relación con los conceptos, Las dos dimensiones típicamente atribuidas a las representaciones conceptuales pueden ser entendidas en términos de información. Por un lado, el contenido intencional es determinado del mismo modo que en las propuestas tradicionales de Fodor y Dretske. El contenido intencional de un concepto *C* está determinado por la relación nomológica que sus instanciaciones tienen con las instanciaciones de una propiedad dada. Por otro lado, el contenido cognitivo de *C* es determinado por los contenidos intencionales, entendidos del mismo modo, de las estructuras que figuran en las inferencias por default en las que aparece *C*. Los contenidos intencionales informativos de las estructuras con las que *C* se relaciona causalmente y que figuran más frecuentemente en las interacciones con los miembros de la clase a la que *C* refiere, constituyen el contenido cognitivo de *C*.

En segundo lugar, la noción de representación perceptual puede entenderse de un modo no semántico, en términos de los sistemas a los que pertenecen. Sin embargo, esta caracterización implica que las representaciones perceptuales tienen un contenido informativo, sólo en el sentido de contenido intencional. Como vimos, la relación de implicación no se da en la otra dirección, no hace falta ser perceptual para tener un contenido informativo intencional. Esto es porque las

representaciones perceptuales son estructuras que tienen la función de responder a tipos de input específicos, a magnitudes físicas específicas. Tienen un contenido informacional intencional porque responden a propiedades distales por medio de responder a magnitudes físicas específicas conectadas causalmente con esas propiedades. Por ejemplo, mi representación visual de la propiedad de ser un perro es una estructura que se activa sólo en la presencia de un patrón de ondas de luz producido por un perro. El hecho de que sólo se active en presencia de un patrón de ondas de luz la hace una representación específicamente visual y, de modo general, perceptual. Ahora bien, si una estructura es activada no sólo por la presencia de un patrón de ondas de luz producido por un perro sino también por el patrón de movimiento molecular causado por el ladrido de un perro, al no responder a un input específico, no podemos decir que pertenezca a un sistema perceptual, aunque tenga contenido informacional.

Por último, el conocimiento es adecuadamente caracterizado en términos de información doxástica. El conocimiento es una creencia que porta la información que tiene la función de portar. Esta condición es equivalente a la condición de seguridad de Pritchard, que he defendido con argumentos propios en contra de la alternativa de las epistemologías de la virtud. La propuesta informacional rival de Dretske adolece de los mismos problemas que la epistemología de la virtud. Entender al conocimiento en términos de información y no meramente seguridad nos permite ver la relación conceptual que hay entre dos aspectos del conocimiento, el contenido y la justificación. Estos dos aspectos se relacionan como una función y el cumplimiento de la misma.

Mostré que los vínculos que tienen estas tres nociones elucidadas con la de información permiten determinar tesis sobre los vínculos conceptuales entre ellas mismas. En primer lugar, defendí un anti-empirismo de conceptos. La tesis de que ser una representación conceptual implica no ser una representación perceptual. Así, a pesar de que tanto las representaciones conceptuales como las perceptuales tienen un contenido caracterizado en términos informacionales, la adquisición del contenido informacional cognitivo propio de los conceptos requiere la intervención de un sistema no perceptual. Esto, como vimos, implica que las representaciones que tienen este tipo de contenido no pueden ser llamadas propiamente perceptuales.

Vimos que la tesis empirista de que toda representación es perceptual supone la noción de copia. Todos los sistemas, según este enfoque, emplean representaciones perceptuales o copias suyas. Vimos que hay dos modos en que podemos entender que estas representaciones, a pesar de que sean

empleadas por otros sistemas, son aún perceptuales, sólo pertenecen a sistemas perceptuales. Un modo de pertenecer a un sistema perceptual es que dichas representaciones hayan sido empleadas originalmente por sistemas perceptuales. Otra opción es que hayan sido *producidas* por dichos sistemas. Tanto ser empleado originalmente como ser producido por un sistema parecen ser criterios razonables de pertenencia a un sistema.

Vimos que según este criterio se sigue que aún si las representaciones perceptuales pueden adquirir contenido conceptual, es decir, pueden tener un contenido cognitivo tal como lo he caracterizado, además de su contenido intencional, dejan de ser perceptuales en el momento en que lo hacen. Pues el contenido informacional cognitivo de una representación *C* es constituido por la formación de relaciones causales entre *C* y cualquier otra representación. En particular, puede siempre involucrar relaciones entre *C* y representaciones que no pertenezcan al mismo sistema perceptual de *C*. Pero los sistemas perceptuales sólo manipulan representaciones que le son propias, no manipulan representaciones que pertenezcan a diferentes sistemas. Por lo tanto, el sistema que vincula a las representaciones que constituyen el contenido cognitivo de un concepto no es un sistema perceptual. Los conceptos son producidos por sistemas no perceptuales y, en este sentido, no son perceptuales. De este modo, las nociones informacionales de contenido conceptual y perceptual defendidas implican una forma de anti empirismo.

Por último, hay un vínculo semejante entre las nociones de conocimiento y percepción. La noción informacional de conocimiento propuesta implica que no hay conocimiento perceptual. El enfoque propuesto sobre el conocimiento implica que las facultades y procesos cognitivos son epistémicamente irrelevantes. En contra de lo que afirman el confiabilismo de procesos y la epistemología de la virtud, las propiedades de estos procesos y facultades no forman parte de las condiciones necesarias para que una creencia constituya conocimiento. Pero parece ser que sólo tiene sentido distinguir tipos de conocimiento si las diferentes facultades fuesen relevantes en este sentido. El conocimiento perceptual, por ejemplo, tiene lugar cuando alguna propiedad de la percepción es parte de lo que hace que una creencia dada sea conocimiento. Si, como he argumentado, esto no ocurre nunca, entonces no hay propiamente tipos de conocimiento y, en particular, no hay conocimiento perceptual.

Hasta aquí repasamos la caracterización informacional de tres tipos de estados centrales en nuestra vida mental y las relaciones que dicha caracterización nos permite establecer entre ellos. En la segunda parte he profundizado en las relaciones que guardan estos estados así entendidos con las

propiedades más básicas en las que se supone que se realizan. Como vimos, esta tarea es un componente medular de la defensa de un enfoque informacional sobre las representaciones. Pues la ventaja central que se supone que tiene esta teoría es que hace a la psicología intencional compatible con el naturalismo, ofrece condiciones físicas para la individuación de los contenidos mentales. Debemos mostrar que esto es efectivamente así, que el enfoque informacional nos muestra cómo la cognición se realiza sobre una base computacional y neuronal.

En primer lugar, me ocupé, no de la cognición, sino de los tipos de procesos que constituyen su base de realización: La computación y el procesamiento neuronal. Presenté y adopté la caracterización que Piccinini y Bahar (2013) ofrecen de la noción de computación. Piccinini y Bahar distinguen la noción de computación de la noción de información con la que es comúnmente confundida. Mostré, sin embargo, que la distinción entre información y computación es más fuerte que la que sostienen estos autores. No es sólo el caso de que la noción de computación no implica la de información sino que la de información tampoco implica computación. A su vez, mostré que esta noción de computación implica una relación más estrecha entre procesamiento neuronal y computación que la que comúnmente se cree. El procesamiento neuronal no constituye una base de realización para la computación sino que es simplemente un tipo específico de computación, junto a la digital y la analógica.

En el capítulo 8 comencé a relacionar a la cognición con su base inmediata computacional. En particular, evalué la tesis fodoriana de que la cognición, informacionalmente caracterizada, no implica ni es implicada por la computación a la luz de los desarrollos recientes discutidos en el capítulo 7. Sostuve que la noción fodoriana de computación, en términos de manipulación de formas sintácticas, debe ser abandonada a favor de la de Piccinini y Bahar, pero que, en contra de lo que implican tesis de estos autores, su noción no es implicada por la de cognición, informacionalmente entendida. Además, afirmé que si bien la noción de forma sintáctica no nos sirve para caracterizar a la computación, sí nos sirve para caracterizar aspectos importantes de la noción informacional de cognición. La cognición sí implica la manipulación de formas sintácticas y esta característica pone restricciones relevantes para la tesis de que la cognición es un proceso computacional.

Por último, en el capítulo 9, me ocupé de la relación de la cognición con el procesamiento neuronal. Argumenté que aunque no hay una relación de implicación de computación a cognición, sí puede haberla de procesamiento neuronal a cognición, si entendemos al procesamiento neuronal no

meramente en virtud de los vehículos que manipula sino también a partir de la función del sistema que lo lleva a cabo. Antes de argumentar a favor de esta relación de implicación, sin embargo, defendí una caracterización peculiar de la noción de retroalimentación neuronal negativa, en el contexto del debate que Wimsatt (1971) llevo adelante en torno a esta noción.

Por último, en el capítulo 10, me introduje en el debate contemporáneo en torno a la noción de realización para determinar qué nos dicen las relaciones conceptuales establecidas en los tres capítulos anteriores sobre las relaciones metafísicas que mantienen la cognición, la computación y el procesamiento neuronal. Discutí los dos enfoques más difundidos actualmente: el enfoque plano y el dimensionado. En particular, argumenté que si bien debemos abandonar el enfoque dimensionado, por ser incompatible con la tesis del fisicalismo de casos, el enfoque más adecuado carece de varios de los supuestos constitutivos del enfoque plano. De estas tesis resultó una especie de enfoque híbrido a partir de cual pudimos determinar que las nociones relevantes estudiadas en los capítulos anteriores mantienen tres relaciones diferentes entre sí. La computación efectivamente no es realizada por el procesamiento neuronal sino que este es simplemente un tipo específico de aquella. La cognición, por su parte, sí se realiza en procesos computacionales y neuronales, pero de un modo diferente en cada uno de ellos. Se realiza de un modo metafísicamente necesario en el procesamiento neuronal (la instanciación del procesamiento neuronal es metafísicamente suficiente para la instanciación de un proceso cognitivo o informacional) y se realiza de un modo metafísicamente contingente sobre la computación (la instanciación de un proceso computacional es sólo nomológicamente suficiente para la instanciación de un proceso cognitivo o informacional). Estas dos últimas tesis eran justamente las que necesitamos para defender un enfoque informacional de la cognición. Tenemos ahora un suelo firme, desde el punto de vista naturalista, para adoptar un enfoque informacional porque podemos explicar de qué modo este enfoque permite la naturalización del contenido, podemos explicar por qué los procesos informacionales pueden ser idénticos a procesos de niveles inferiores, físicos, neuronales o computacionales.

Con esto finaliza al menos una primera versión de la defensa del enfoque informacional de la mente. He ofrecido diferentes tipos de razones para pensar que los contenidos de las representaciones perceptuales y de las perceptuales y el conocimiento son adecuadamente caracterizados en términos informacionales. Mostré que las intuiciones y argumentos detrás de una caracterización adecuada de estos tipos de estados cuadran con una concepción informacional de los mismos. Además, mostré que los contenidos mentales así entendidos son aceptables desde el punto de vista científico, pueden ser

algo más que meras nociones del sentido común, científicamente prescindibles o cuyo abandono podemos recomendar, sino que pueden ser integradas a una visión científica del mundo en la medida en que podemos dar una explicación clara de la relación que mantienen con propiedades físicas de nivel más básico.

Referencias

Adams, F., Aizawa, K. 2010. "Causal Theories of Mental Content", en Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. URL =

<<http://plato.stanford.edu/archives/spr2010/entries/content-causal/>>

Adrian, E. D. 1928. *The Basis of Sensation: The Action of the Sense Organs*. New York: Norton.

Aizawa, K. & Gillett, C. 2009a. "The (Multiple) Realization of Psychological and other Properties in the Sciences", *Mind & Language*, 24(2), pp. 181-208.

Aizawa, K. & Gillett, C. 2009b. "Levels, Individual Variation, and Massive Multiple Realization in Neurobiology", en Bickle (comp.) *Oxford Handbook of Philosophy and Neuroscience*, Oxford University Press.

Ashby, W. R., 1960. *Design for a Brain*. New York. Wiley (segunda edición revisada).

Axtell, G. 1997. 'Recent Work in Virtue Epistemology', *American Philosophical Quarterly* 34, 410-30.

Barberis, S. 2014. "Lo que la realización no es", de próxima publicación como capítulo en una compilación editada por Liza Skidelsky, Buenos Aires, *Agencia Nacional de Promoción Científica y Tecnológica*.

Baddeley, R., Hancock, P., et al. (eds.) 2000. *Information Theory and the Brain*. Cambridge. Cambridge University Press.

Beckner, M., 1959. *The Biological Way of Thought*. New York. Columbia Univ. Press.

Baehr, J. 2004. "Virtue Epistemology", *Internet Encyclopaedia of Philosophy*, (eds.) B. Dowden & J. Fieser, URL= www.iep.utm.edu/virtueep/.

Bertalanfy, L. von, 1950 'The Theory of Open Systems in Physics and Biology', *Science* 111, 23-29.

Bartalanfy, L. von, 1950. "An Outline of General Systems Theory", *Brit. J. Phil. Sci.* 1, 139-64.

Biederman, I. 1987. "Recognition-by-Components: A theory of Human Image Understanding". *Psychological Review* 94: 115-147

Block, N. 1978. "Troubles with Functionalism" en *Perception and Cognition: Issues in the Foundation of Psychology*. W. C. Savage, (ed.), Minneapolis. University of Minnesota Press.

Braithwaite, R. B., 1953. *Scientific Explanation*. London. Cambridge University Press. Capítulo 10.

Cannon, W. B. 1932. *The Wisdom of the Body*. New York. Norton.

Chisholm, R. 1977. *Theory of Knowledge*, (segunda edición), Englewood Cliffs, NJ. Prentice-Hall.

Churchland, P. M. 1988. "Perceptual Plasticity and Theoretical Neutrality: A Reply to Jerry Fodor", *Philosophy of Science* 55: 167-187

Churchland, P. S., & Sejnowski, T. J. 1992. *The computational brain*. Cambridge, MA. MIT Press.

Craver, C. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*, Oxford. Clarendon Press.

Craver, C. .2009. "Mechanisms and Natural Kinds", *Philosophical Psychology* 22, pp. 575-594.

Craver, C. F. 2010. "Prosthetic Models", *Philosophy of Science* 77(5), 840-851.

Craver, C. 2013. "Functions and Mechanisms: A Perspectivalist View", en P. Huneman (Ed.) *Functions: Selection and Mechanisms*, Dordrecht. Springer, pp. 133-158.

Craver, C. y Bechtel, W. 2007. "Top-Down Causation Without Top-Down Causes", *Biology and Philosophy*, 22, pp. 547-563.

Dayan, P. and L. F. Abbott 2001. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA. MIT Press.

Devitt, M. 1981. *Designation*. New York. Columbia University Press

Donnellan, K. S. 1972. "Proper Names and Identifying Descriptions" En *Semantics of Natural Language*. D. Davidson and G. Harman (editores). Dordrecht. Reidel.

Dretske, F. I. 1981. *Knowledge and the Flow of Information*. Cambridge. MIT Press.

Edelman, S. 2008. *Computing the mind: How the mind really works*. Oxford, England. Oxford University Press.

Endicott, R. 2011. “Flat Versus Dimensioned: the What and the How of Functional Realization”, *Journal of Philosophical Research*, 36, pp. 191-208.

Ermentrout, G. B. and Terman, D. H. 2010. *Mathematical Foundations of Neuroscience*. New York: Springer.

Fodor, J. A. 1990. *A Theory of Content and Other Essays*. Cambridge. MIT Press.

Fodor, J. A. 1991. “A Modal Argument for Narrow Content”, *Journal of Philosophy* 88: 5–25.

Fodor, J. A. with E. Lepore. 1992. *Holism: A Shopper’s Guide*. Oxford. Basil Blackwell.

Fodor, J. A. 1983. *The Modularity of Mind*. Cambridge. MIT Press

Fodor, J. A. 1994. *The Elm and the Expert: Mentalese and Its Semantics*. Cambridge. MIT Press.

Fodor, J. A. 1998. *Concepts: Where Cognitive Science Went Wrong*, The 1996 John Locke Lectures, Oxford University Press

Fodor, J. A. 2000. *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind*. Cambridge. MIT Press.

Fodor, J. A. 2008. *LOT 2, The Language of Thought Revisited*. Oxford. Oxford University Press

Fodor, J. A., Lepore, E. 1992. *Holism: A Shopper's Guide*. Oxford. Basil Blackwell,

Fodor, J. A., Lepore, E. 2002. *The compositionality Papers*. Oxford. Oxford University Press.

Frege, G. "Über Sinn und Bedeutung", en *Zeitschrift für Philosophie und philosophische Kritik* 100: 25–50, 1892. Trad. de M. Black, Oxford: Blackwell, 1953.

Garson, J. 2003. "The Introduction of Information into Neurobiology", *Philosophy of Science* 70, 926-936.

Galistel, C. R., & King, A. P. 2009. *Memory and the computational brain: Why cognitive science will transform neuroscience*. Malden, MA. Wiley-Blackwell.

Gettier, E. 1963. "Is Justified True Belief Knowledge?", *Analysis* 23, 121-23.

Gillett, C. 2002. "The Dimensions of Realization", *Analysis*, 62, pp. 316-323.

Gillett, C. 2003. "The Metaphysics of Realization, Multiple Realizability and the Special Sciences", *Journal of Philosophy*, 100, pp. 591-603.

Gillett, C. 2007. "Understanding the New Reductionism: The Metaphysics of Science and Compositional Reduction", *Journal of Philosophy* 104(4), pp. 193-216.

Goldberg, Sanford C. 2005. "Testimonial Knowledge through Unsafe Testimony," *Analysis* 65: 302-11.

Goldberg, S. 2010. *Relying on others, An essay in epistemology*. Oxford. Oxford University Press.

Goldman, A. 1976. "Discrimination and Perceptual Knowledge", *Journal of Philosophy* 73, 771-91.

Greco, J. 2009c. "Knowledge and Success From Ability", *Philosophical Studies* 142, 17-26.

Greco, J. 2002. "Virtues in Epistemology", *Oxford Handbook of Epistemology*, P. Moser (editor), 287-315, Oxford. Oxford University Press.

Greco, J., & Turri, J. 2009. "Virtue Epistemology", *Stanford Encyclopaedia of Philosophy*, E. Zalta (editor), URL= <http://plato.stanford.edu/entries/epistemology-virtue/>.

Grice, H. P. 1957. "Meaning", *The Philosophical Review*, 66(3), 377-388.

Kacser, H., 1957. "Some Physico-Chemical Aspects of Biological Organization", impreso como apéndice a Waddington, C. H. 1957. *The Strategy of the Genes*, New York. Macmillan, pp. 191-249.

Kim, J. 1998. *Mind in a Physical World*, Cambridge, MIT Press.

Kleene, S. C., 1956 “Representation of Events in Nerve Nets and Finite Automata” en *Automata Studies*, C. E. Shannon and J. McCarthy (editores), Princeton, N. J., Princeton Univ. Press, pp. 3-42.

Kossylyn, S. M. 1980. *Image and Mind*. Cambridge. Harvard University Press.

Kripke, S. A. 1980. *Naming and Necessity*. Cambridge. Harvard University Press.

Kvanvig, J. 2010. “Virtue Epistemology”, en *Routledge Companion to Epistemology*, S. Bernecker & D. H. Pritchard (editores), London. Routledge. pp. 199-207.

Lackey, J. 2008. *Knowing from words*. Oxford. Oxford University Press.

Lin, J.-W. and Faber D. S. 2002. “Modulation of Synaptic Delay During Synaptic Plasticity”, *Trends in Neurosciences* 25(9), 449-455.

Luper, S. 1984. “The Epistemic Predicament”, *Australasian Journal of Philosophy* 62, 26-50.

Lewis, D. 1970 “How to Define Theoretical Terms”, *Journal of Philosophy* 77: 427-446.

Lewis, D. 1970. “How to Define Theoretical Terms”, *Journal of Philosophy* 67: 427-446

Marr, D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco. W. H. Freeman.

McCulloch, W. S. y Pitts W. H. 1943. “A Logical Calculus of the Ideas Immanent in Nervous Activity”, *Bulletin of Mathematical Biophysics* 7, 115-133.

Mills, J. W. 2008. “The Nature of the Extended Analog Computer”. *Physica D: Nonlinear Phenomena*, 237(9), 1235-1256.

Nagel, E., 1961. *The Structure of Science*. New York. Harcourt.

Newell, A. 1990. *Unified theories of cognition*. Cambridge, MA. Harvard University Press.

Newell, A., & Simon, H. A. 1976. “Computer science as an empirical enquiry: Symbols and search”, *Communications of the Association for Computing Machinery*, 19, 113–126.

Piccinini, G. 2007. “Computing Mechanisms”, *Philosophy of Science* 74(4), 501-526.

Piccinini, G. 2004. “The First Computational Theory of Mind and Brain: A Close Look at McCulloch and Pitts’s ‘Logical Calculus of Ideas Immanent in Nervous Activity’”, *Synthese* 141(2), 175-215.

Piccinini, G. and Scarantino, A. 2011. “Information processing, Computation and Cognition”, *Journal of Biological Physics*, 37.1, 1-38.

Piccinini, G. and Bahar, S. 2013. “Neural Computation and the Computational Theory of Cognition”, *Cognitive Science*, 37-3, pp. 453-488.

Polger, T. 2004. *Natural Minds*, Cambridge. MIT Press.

Polger, T. y Shapiro L. 2008. "Understanding the Dimensions of Realization", *Journal of Philosophy*, 105(4), pp. 213-222.

Pour-El, M. B. 1974. "Abstract Computability and Its Relation to the General Purpose Analog Computer (Some Connections Between Logic, Differential Equations and Analog Computers)", *Transactions of the American Mathematical Society*, 199, pp. 1-28.

Putnam, H. 1967. "The Nature of Mental States" en H. Putnam (1975) *Mind, Language and Reality: Philosophical Papers Vol. 2*, New York. Cambridge University Press, pp. 429-440.

Putnam, H. 1975. "The Meaning of 'Meaning' " en H. Putnam, *Philosophical Papers*, vol. 2: *Mind, Language, and Reality*. Cambridge. Cambridge University Press.

Prinz, J. 2002. *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA. MIT Press.

Prinz, J. 2005. "The Return of Concept Empiricism," En *Categorization and Cognitive Science*, H. Cohen and C. Lefebvre (Editores), Elsevier, pp. 680-695.

Prinz J. 2006. "Beyond Appearances: The Content of Perception and Sensation," en *Perceptual Experience*, T. S. Gendler y J. Hawthorne (Editores), Oxford University Press, pp. 434-459.

Pritchard, D. H. 2002. "Resurrecting the Moorean Response to the Sceptic", *International Journal of Philosophical Studies* 10, 283-307.

Pritchard, D. H. 2003. "Virtue Epistemology and Epistemic Luck", *Metaphilosophy* 34, 106-30.

Pritchard, D. H. 2005. *Epistemic Luck*, Oxford. Oxford University Press.

Pritchard, D. H. 2007a. "Anti-Luck Epistemology", *Synthese* 158, 277-97.

Pritchard, D. H. 2013. "The Power, and Limitations, of Virtue Epistemology" (con J. Kallestrup), en *Powers and Capacities in Philosophy: The New Aristotelianism*, J. Greco y R. Groff (editores), Routledge.

Pylyshyn, Z. W. 1984. *Computation and cognition*. Cambridge, MA. MIT Press.

Quine, V. O. 1953. "Two Dogmas of Empiricism" en *From a Logical Point of View*, Cambridge, Mass. Harvard University Press, pp. 20-46.

Sainsbury, R. M. 1997. "Easy Possibilities", *Philosophy and Phenomenological Research* 57, 907-19.

Schneider, S. 2011. *Language of thought: A new philosophical direction*. Cambridge, MA. MIT Press.

Shapiro, L. 2000. "Multiple Realizations", *Journal of Philosophy*, 97(12), pp. 635-654.

Shapiro, L. 2004. *The Mind Incarnate*, Cambridge. MIT Press.

Shoemaker, S. 2007. *Physical Realization*, Oxford. Oxford University Press

Smart, J. J. .1959. "Sensations and Brain Processes", *The Philosophical Review*, 68(2), pp. 141-156.

Stampe, D. 1979. "Toward a Causal Theory of Linguistic Representation" en *Contemporary Perspectives in the Philosophy of Language*, P. A. French, T. E. Uehling Jr., y H. K. Wettstein (editores), *Midwest Studies in Philosophy*, no. 2. Minneapolis. University of Minnesota Press.

Stoljar, D. 2009. "Physicalism", en E. Zalta (Editor) *The Stanford Encyclopedia of Philosophy*, URL=<http://plato.stanford.edu/archives/fall2009/entries/physicalism/>>.

Sosa, E. 1999. "How to Defeat Opposition to Moore", *Philosophical Perspectives* 13, 141-54.

Stich, S. P. 1983. *From Folk Psychology to Cognitive Science: The case against Belief*. Cambridge. MIT Press.

Sutherland, N. S. 1968. "Outline of a Theory of Visual Patter Recognition in Animals and Man", *Proceedings of the Royal Society of London B* 171: 297-317.

Turri, J., & Sosa, E. 2010. "Virtue Epistemology", *Oxford Bibliographies: Philosophy*, DOI: 10.1093/OBO/9780195396577-0123.

Wajnerman Paz, A. 2013. "Modal considerations on computation and information processing" en *Journal of Cognitive Science*, Korea. Institute for Cognitive Science (Seoul National University), Korea, pp. 253-286.

Wiener, N. 1948. *Cybernetics or control and communication in the animal and the machine*. Cambridge, MA. MIT Press.

Wimsatt, W. 1971. "Some Problems with the Concept of Feedback", en *Boston Studies in the Philosophy of Science*, volume 8, R. C. Buck and R. S. Cohen (editores), Dordrecht. Reidel, pp. 241-256.

Wimsatt, W. 1994. "The Ontology of Complex Systems: Levels, Perspectives and Causal Thickets", *Canadian Journal of Philosophy*, 20, pp. 207-274.

Williamson, T. 2000. *Knowledge and its Limits*, Oxford. Oxford University Press.

