



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

P

En defensa de una teoría pluralista de conceptos

Autor:

Haimovici, Sabrina

Tutor:

Skidelsky, Liza

2015

Tesis presentada con el fin de cumplimentar con los requisitos finales para la obtención del título Doctor 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

UNIVERSIDAD DE BUENOS AIRES

Facultad de Filosofía y Letras

**En defensa de una teoría pluralista de
conceptos**

Tesis de Doctorado

2015

Doctoranda: Lic. Sabrina Haimovici

Directora: Dra. Liza Skidelsky

Índice

Agradecimientos	5
Introducción	6
Capítulo 1. Requisitos explicativos de una teoría de conceptos	21
1. Requisitos relacionados con el poder expresivo del pensamiento	21
2. Requisitos relacionados con el uso de los conceptos en procesos cognitivos	27
3. Requisitos fundamentales de una teoría de conceptos	35
Capítulo 2. Tipos de estructura interna. Primera parte: definiciones, prototipos y ejemplares	42
1. La teoría clásica de las definiciones	42
1.1 Las ventajas de la teoría de las definiciones	43
1.2 Las objeciones a la teoría de las definiciones	44
2. El enfoque de prototipos	47
2.1 Ventajas y evidencia a favor de los prototipos	49
2.2 Las dificultades del enfoque de prototipos	52
3. El enfoque de ejemplares	55
3.1 Ventajas y evidencia a favor de los ejemplares	57
3.2 Dificultades de este enfoque	58
4. La estructura semántica interna y la individuación del contenido	59
5. Las ventajas de combinar definiciones, prototipos y ejemplares con otros tipos de estructuras	61
Capítulo 3. Tipos de estructura interna. Segunda parte: las teorías	65
1. El enfoque de teoría-teoría de los conceptos	66
2. Los conceptos como partes de teorías	70
2.1 Los vínculos entre los conceptos y las estructuras más amplias de conocimiento....	71
2.2 Evidencia a favor de los vínculos constitutivos entre conceptos y teorías	75
3. Los conceptos como teorías	78
3.1 Estructuras de rasgos relacionados causalmente	78
3.2 Evidencia a favor de la estructura interna análoga a estructuras teóricas.....	79
4. Dificultades del enfoque de teoría-teoría.....	85
5. La integración de teorías en una propuesta pluralista	88
Capítulo 4. El atomismo conceptual	91
1. Las tesis principales del atomismo informacional	91
2. Las ventajas del atomismo	97

2.1 Alcance	97
2.2 Composicionalidad	98
2.3 Intencionalidad	100
3. Las dificultades del atomismo.....	101
3.1 La individuación de conceptos coextensivos	101
3.2 Adquisición	107
3.3 Categorización.....	110
Capítulo 5. El neo-empirismo de conceptos.....	120
1. El neo-empirismo de conceptos	121
2. La evidencia y las ventajas del neo-empirismo	125
2.1 La evidencia a favor de los vehículos modales	125
2.2 Categorización.....	129
2.3 Adquisición	130
3. Las dificultades del neo-empirismo	132
3.1 Intencionalidad	133
3.2 Alcance	138
3.3 Composicionalidad	145
Capítulo 6. Las teorías híbridas, pluralistas y eliminativitas	153
1. Propuestas híbridas y pluralistas	154
2. Las condiciones de coordinación y asociación	155
2.1 La condición de coordinación	157
2.2 La condición de asociación	162
3. El dilema de Weiskopf	168
4. Algunas estrategias para rechazar el eliminativismo	169
Capítulo 7. El pluralismo amplio	178
1. La necesidad de postular un pluralismo amplio a partir de los requisitos explicativos fundamentales	179
2. El poder explicativo del pluralismo amplio	185
3. Tipos de vehículos conceptuales y Tipos de procesamiento	205
4. El pluralismo amplio y una arquitectura dual	212
5. El caso de los conceptos de números	214
6. El pluralismo amplio y las ventajas de una arquitectura dual.....	219
Conclusiones.....	226

Referencias 234

Agradecimientos

Quisiera expresar, en primer lugar, mi mayor agradecimiento a Liza Skidelsky, por su enseñanza y guía en la investigación filosófica. Su generosidad y dedicación al discutir mis ideas y desafiarme a profundizar permitieron que una pequeña intuición se desarrollara y articulara en una tesis.

Quiero agradecer al grupo de investigación Cognición, Lenguaje y Percepción, coordinado por Liza Skidelsky, por ofrecerme el espacio ideal para pensar y discutir, en el que tomaron forma muchas de las ideas que desarrollo en esta tesis. Agradezco especialmente a Sergio Barberis, Mariela Destéfano, Abel Wajnerman, Nicolás Serrano y Fernanda Velázquez, por compartir la idea de que la investigación filosófica no es una tarea individual.

Quiero agradecer también a mi familia, por el apoyo que me brindaron. A mis hermanos Julia y Ariel, por acompañarme siempre, especialmente en el último tiempo de escritura de tesis. A Patricio, mi marido, por estar conmigo, por tenerme paciencia, porque con él soy feliz.

Por último, quisiera dedicar esta tesis a mi mamá, Daniela, y a mi papá, Luciano. Ninguna de mis ideas sería posible si yo no hubiese sido, en algún momento, una idea de ellos. Gracias por todo lo que me dieron desde esa idea hasta hoy.

Introducción

Entre las capacidades cognitivas que exhibe la mente humana, nuestra capacidad para pensar acerca de categorías ocupa un lugar central en las investigaciones de la ciencia cognitiva. Podemos tener pensamientos acerca de lo que nos rodea y agrupar distintas instancias como pertenecientes a una misma categoría. Podemos, por ejemplo, pensar en los perros en general o en uno en particular. Podemos reconocer instancias de categorías y, en consecuencia, atribuirles determinadas propiedades. Por ejemplo, ante una manzana, pensamos que es comestible, esperamos que tenga cierta textura, cierta acidez y cierto dulzor. No necesitamos investigar todas sus características desde cero, la reconocemos como manzana. Podemos pensar acerca de lo que percibimos o acerca de entidades que no están presentes en el entorno inmediato. También podemos combinar distintos pensamientos para generar otros nuevos. Así, podemos planear eventos, por ejemplo un viaje, pensar que en unos meses estaremos en una ciudad nueva, hacer una lista imaginaria de los lugares que queremos recorrer, de lo que queremos llevar en la valija.

Podemos pensar acerca de entidades o propiedades de distinto tipo. Podemos pensar en propiedades concretas y abstractas, por ejemplo, en el color verde o en la injusticia. Podemos pensar acerca de clases biológicas, como la clase de los rinocerontes, o artificiales, como la clase de las impresoras. También podemos pensar en entidades ficticias o mitológicas, como Sherlock Holmes o los minotauros. Podemos tener pensamientos acerca de categorías nuevas, en las que nunca antes habíamos pensado. Por ejemplo, podemos aprender que los *patisson* son una variedad de *cucurbita* (la especie a la que pertenecen los zapallitos). Todas estas capacidades están relacionadas con nuestros conceptos. Los conceptos son los constituyentes de los pensamientos, nos permiten tratar distintas instancias como pertenecientes a una misma categoría, proporcionan un conjunto de elementos mínimos que se pueden combinar para formar pensamientos complejos y nos permiten realizar inferencias.

En los últimos años, en los ámbitos de la ciencia cognitiva y la filosofía de la mente, se han elaborado distintas teorías respecto de qué son los conceptos, qué estructura tienen, cómo se adquieren y cómo intervienen en diversos procesos mentales. En general, hay algo en lo que todas acuerdan: los conceptos son entidades encargadas de explicar gran parte de las características de nuestra vida mental y de la relación entre nuestras mentes y el mundo. Hay dos metáforas que suelen asociarse a los conceptos, cada una basada en una de sus funciones principales. Por un lado, se los caracteriza como los “bloques para la construcción de pensamientos”, aludiendo a que son los constituyentes de los pensamientos, las unidades

mínimas que se combinan para formarlos. En este sentido, los conceptos serían como bloques del juego LEGO, que se pueden ensamblar para armar construcciones. Algunos autores, como Fodor (1987), señalan que dado que los conceptos son los constituyentes de los pensamientos, estos son las entidades que permiten explicar, por un lado, la intencionalidad de los pensamientos y, de esta manera, la relación entre los pensamientos y el mundo y, por otro lado, la realización de inferencias deductivas que dependen de la estructura de los pensamientos.

La segunda metáfora compara los conceptos con una suerte de “pegamento” que otorga cohesión a nuestra vida mental. Esta segunda metáfora enfatiza el rol de los conceptos como portadores de conocimiento acerca de categorías y el uso que hacemos de ellos, especialmente en las tareas de categorización (Murphy 2002, Smith & Medin 1981). El hecho de identificar distintas instancias como correspondientes a una misma categoría, atribuyéndoles el mismo concepto, organiza en parte nuestro pensamiento. Estructuramos el mundo que nos rodea en categorías, agrupando diversas instancias bajo cada una de ellas. Más allá de si estas categorías son clases naturales o no, es decir, si recortamos el mundo “por sus articulaciones” o nuestras categorías son construcciones culturales proyectadas sobre él, nuestro comportamiento está guiado en parte por los resultados de nuestras categorizaciones. Habitualmente, nos encontramos con instancias nuevas de categorías conocidas y nos comportamos con respecto a ellas atribuyéndoles las características que asociamos a esas categorías, en lugar de investigar sus propiedades en cada encuentro. Cuando entramos en un bar, por ejemplo, encontramos sillas, mesas, probablemente una barra. Sabemos que en las sillas nos podemos sentar, que lo que pidamos lo van a apoyar sobre la mesa. Todo el tiempo realizamos inferencias respecto a las cosas que nos rodean en función de las categorías que les atribuimos. Si probamos un café, esperamos que sea bebible, de color oscuro, que tenga un sabor y un aroma particulares. Anticipamos estas propiedades a partir de la información que tenemos asociada al concepto CAFÉ.¹

Smith y Medin (1981) señalan que sin la capacidad para categorizar, si identificáramos a cada entidad como única, nuestra vida mental resultaría caótica. Su descripción de cómo sería una vida mental sin conceptos recuerda a las características de Funes, el personaje de Borges, que tiene una memoria infinita y recuerda cada instante como único. Parece adecuado describir a Funes como un individuo que carece de categorías generales y representa cada objeto, evento o individuo como único, es decir, como

¹ Utilizaré mayúsculas para los nombres de los conceptos e itálicas para los nombres de los valores semánticos de los conceptos.

perteneciente a una categoría de un único elemento. En palabras de Borges (1974):

No sólo le costaba comprender que el símbolo genérico *perro* abarcara tantos individuos dispares de diversos tamaños y diversa forma; le molestaba que el perro de las tres y catorce (visto de perfil) tuviera el mismo nombre que el perro de las tres y cuarto (visto de frente). [...] Había aprendido sin esfuerzo el inglés, el francés, el portugués, el latín. Sospecho, sin embargo, que no era muy capaz de pensar. Pensar es olvidar diferencias, es generalizar, abstraer. En el abarrotado mundo de Funes no había sino detalles, casi inmediatos. (Borges 1974, p. 490)

Tratar de imaginar una vida mental sin conceptos resulta, aun con ayuda de Borges, sumamente difícil. Es una vida mental muy distinta a la que conocemos. La descripción de Funes resulta interesante porque parece faltarle las características destacadas en las dos metáforas asociadas a los conceptos. En su memoria infinita no figuran representaciones de categorías que permitan conectar las distintas experiencias. Le molesta el uso del mismo símbolo para identificar distintos encuentros con el mismo perro, atribuyéndole a cada evento una categoría específica y perdiendo así la posibilidad de ligar una experiencia con la otra. En este sentido, se podría decir, en términos de Smith y Medin, que el mundo de Funes carece de un “pegamento” que permita ligar un pensamiento con otro, dado que cada pensamiento es único y no comparte categorías generales con otros pensamientos. Así, Funes también carece de los “ladrillos” mínimos para construir pensamientos, puesto que sus distintos pensamientos no tienen elementos en común. El pensamiento de Funes no es productivo ni sistemático, i.e. no consiste en un conjunto finito de elementos que se pueden combinar para construir pensamientos nuevos, ni guarda relaciones estructurales entre distintos pensamientos, sino que dispone de infinitos elementos y cada pensamiento nuevo es completamente distinto del anterior, sin compartir constituyentes uno con otro. Borges (1974) sugiere esta última característica al describir cómo Funes había desarrollado un nuevo sistema numérico que no es productivo ni sistemático:

Me dijo que hacia 1886 había discurrido un sistema original de numeración y que en muy pocos días había rebasado el veinticuatro mil. No lo había escrito, porque lo pensado una sola vez ya no podía borrarse. Su primer estímulo, creo, fue el desagrado de que los treinta y tres orientales requirieran dos signos

y tres palabras, en lugar de una sola palabra y un solo signo. Aplicó luego ese disparatado principio a los otros números. En lugar de siete mil trece, decía (por ejemplo) *Máximo Pérez*; en lugar de siete mil catorce, *El Ferrocarril*; otros números eran *Luis Melián Lafinur*, *Olimar*, *azufre*, *los bastos*, *la ballena*, *el gas*, *la caldera*, *Napoléon*, *Agustín de Vedía*. En lugar de quinientos, decía *nueve*. Cada palabra tenía un signo particular, una especie de marca; las últimas eran muy complicadas... Yo traté de explicarle que esa rapsodia de voces inconexas era precisamente lo contrario de un sistema de numeración. Le dije que decir 365 era decir tres centenas, seis decenas, cinco unidades: análisis que no existe en los "números" *El Negro Timoteo o manta de carne*. (Borges 1974, p. 489)

Nuestro sistema conceptual no funciona como el de Funes. En un sentido, es más limitado. No disponemos de un conjunto infinito de representaciones que podamos asignar a cada experiencia, a cada instante, a cada encuentro con una instancia de una categoría, sino que contamos con un conjunto finito de representaciones. Sin embargo, nuestra capacidad representacional es infinita, puesto que los conceptos se pueden combinar de maneras novedosas para generar pensamientos nuevos. En este sentido, nuestro sistema conceptual tiene una ventaja respecto del de Funes: puesto que utilizamos el mismo conjunto de conceptos para formar distintos pensamientos, estos pueden tener elementos en común, que conectan un pensamiento con otro.

Así, los conceptos cargan con una importante responsabilidad en nuestra vida mental y, en consecuencia, una teoría de conceptos formará parte de la explicación de cómo está estructurado nuestro pensamiento y de cómo funciona, en general, nuestra mente. Puesto que las funciones que cumplen los conceptos son múltiples y diversas, generalmente las teorías se concentran en la explicación de algunas de ellas. Como consecuencia, muchas de las teorías propuestas parecen adecuadas para dar cuenta de algunos de los fenómenos asociados con los conceptos, pero ninguna consigue acomodar todos los requisitos explicativos que se han propuesto. En esta tesis me ocuparé de evaluar las ventajas y los límites explicativos de las principales teorías de conceptos y desarrollaré una propuesta alternativa con el objetivo de superar las dificultades de las distintas teorías. Sostendré que dados los distintos roles funcionales que los conceptos deben satisfacer, el enfoque adecuado para una teoría de conceptos es de tipo pluralista, i.e. una teoría que postula distintos tipos de estructuras conceptuales. Una propuesta de tipo pluralista permite conservar las ventajas que presenta

cada estructura conceptual y, a la vez, puede desprenderse de las limitaciones que presenta cada estructura por separado.

Se puede clasificar las distintas teorías de conceptos que se han propuesto en función de la postura que adoptan respecto de una serie de debates centrales en torno a los mismos. Estos debates son, particularmente, acerca de la ontología de los conceptos, los requisitos explicativos de una teoría de conceptos, la estructura conceptual interna y el formato o vehículo de los conceptos. En lo que sigue describiré de manera general qué está en juego en los debates centrales, los cuales abordaré a lo largo de la tesis. Siguiendo a Margolis y Laurence (2014), se pueden delinear tres grandes perspectivas en función de las características ontológicas asignadas a los conceptos: la primera sostiene que estos son representaciones mentales, la segunda los identifica con habilidades o capacidades y la última con entidades abstractas. La perspectiva según la cual los conceptos son representaciones mentales postula que estos son entidades psicológicas, esto es, particulares mentales que son los constituyentes básicos del pensamiento. Algunos consideran que se trata de particulares mentales que presentan una semántica composicional y una sintaxis combinatoria semejante a la de los lenguajes naturales (Fodor 1975). Otros autores que adoptan el enfoque de las representaciones mentales postulan, en cambio, que están consisten en copias de representaciones perceptivas (Barsalou 1999, Prinz 2002).

En segundo lugar, quienes consideran que los conceptos son habilidades sostienen que estos consisten en ciertas capacidades cognitivas de los agentes, tales como extraer inferencias o discriminar entre los ítems que caen bajo ellos y aquellos que no. Una de las críticas principales al enfoque de los conceptos como habilidades es que no puede explicar cómo intervienen los conceptos en los procesos mentales. De esta manera, resulta vulnerable a la objeción de que esas mismas habilidades que el enfoque propone se explican mejor apelando a representaciones mentales subyacentes y procesos que operan sobre ellas (Margolis & Laurence 2014). Por último, la postura que entiende los conceptos como sentidos fregeanos los identifica con objetos abstractos. Según esta perspectiva, los conceptos no están “en la cabeza”, sino que son entidades abstractas independientes de la mente, que ésta capta cuando tiene un pensamiento. Las principales críticas a este enfoque señalan, por un lado, que no explica en qué consiste la relación de “captación” que se produce entre nuestras mentes y los sentidos (Margolis & Laurence 2014) y, por otro lado, que no explica por qué captamos cada sentido de un único modo, i.e. por qué no podemos tener distintos vínculos epistémicos con cada sentido extramental (Fodor 1999).

Adoptaré el enfoque de los conceptos como representaciones mentales, dado que es el

que ha resultado más fructífero en psicología y el que resulta más promisorio para integrar una teoría de conceptos con una de la arquitectura cognitiva (Carey 2009, Fodor 1999, Margolis & Laurence 2007, Prinz 2002, *inter alia*). Por esta razón, en la reconstrucción del resto de los debates, i.e. los debates acerca de los requisitos explicativos de las teorías de conceptos, la estructura informacional de los conceptos y los vehículos conceptuales, sólo analizaré las propuestas teóricas que adoptan este enfoque.

Distintas corrientes, tanto en psicología cognitiva como en filosofía de la psicología, apelan a los conceptos para explicar diversos fenómenos como la categorización, el razonamiento inductivo y deductivo, el aprendizaje de relaciones causales, y la sistematicidad y productividad del pensamiento. Se puede agrupar estos fenómenos en virtud de su relación con el poder expresivo del pensamiento y con el uso de los conceptos en procesos cognitivos. Así, las teorías de conceptos pretenden dar cuenta, por un lado, del poder expresivo o capacidad representacional de nuestro pensamiento, intentando ofrecer explicaciones del contenido, i.e. aquello acerca de lo cual son los pensamientos, el alcance, i.e. la amplia gama de conceptos de la que disponemos (que incluye conceptos tales como los de objetos, eventos, conceptos abstractos y teóricos), y la capacidad productiva y sistemática del pensamiento, i.e. el hecho de que podemos pensar infinitos pensamientos y el hecho de que podemos pensar pensamientos relacionados entre sí. Por otro lado, pretenden explicar determinados procesos cognitivos, en particular la adquisición de conceptos (a la que en ocasiones se refieren como aprendizaje de categorías), la categorización (tanto el reconocimiento de ejemplares que caen bajo una categoría cuanto el anticipo de las características que tendrá un ejemplar por pertenecer a una categoría) y el razonamiento.

No todas las teorías toman en cuenta todos estos fenómenos, sino que por lo general se concentran en algunos de ellos. Algunos autores distinguen entre los *desiderata* o requisitos explicativos de una teoría psicológica de conceptos y los de una teoría filosófica. Identifican, a grandes rasgos, las teorías filosóficas como aquellas que se concentran en la explicación del poder expresivo del pensamiento y las teorías psicológicas como aquellas que se encargan de explicar el uso de los conceptos en procesos cognitivos (Fodor 2000b, Machery 2009). Como desarrollaré en el capítulo 1, considero que esta distinción no resulta adecuada, fundamentalmente por dos razones. Primero, esta distinción parece ignorar las relaciones que hay entre los distintos requisitos que, en muchos casos, hacen que sea imposible dar cuenta de uno de ellos sin asumir compromisos (aunque sea de manera implícita) respecto de otros. Segundo, la distinción parece ignorar también la interdisciplinariedad propia de la ciencia cognitiva, por la cual no resulta muy común el

abordaje de los fenómenos cognitivos desde un punto de vista exclusivamente filosófico o psicológico. Las teorías de conceptos generalmente involucran elementos propios de las dos disciplinas. Por otra parte, los debates que se han suscitado entre filósofos y psicólogos en torno a los conceptos han contribuido al desarrollo de distintas propuestas teóricas, lo que sugiere que se ocupan del mismo objeto de estudio. En el capítulo 1 retomaré este debate, desarrollaré cada uno de los requisitos en detalle y seleccionaré un conjunto de requisitos fundamentales que me permitirá evaluar las teorías de conceptos dominantes.

Los otros dos debates principales se dan en torno a la estructura informacional o semántica de los conceptos y a su formato o vehículo. Las representaciones mentales poseen propiedades semánticas y sintácticas o formales. Estas propiedades corresponden, respectivamente, al contenido intencional de las representaciones, i.e. aquello que los conceptos representan, y a su vehículo, i.e. aquello que porta las propiedades semánticas/intencionales (Carey 2009, Camp 2007, Fodor 1999, Prinz 2002). Las propiedades semánticas pueden abordarse en dos dimensiones. Por un lado, la relativa a la relación entre los conceptos y el mundo: acerca de qué son los conceptos, qué tipo de información portan y cómo adquieren ese contenido representacional. Por otro lado, hay una dimensión interna, relativa a la estructura de los conceptos. Estas dos dimensiones están relacionadas y generalmente se las trata conjuntamente, en tanto las tesis relativas a una de ellas conllevan compromisos respecto de la otra. A continuación presentaré brevemente los debates en torno a las propiedades semánticas tomando como guía la dimensión de la estructura interna.

Los primeros debates en torno a qué estructura deben tener los conceptos estuvieron guiados en parte por la aceptación del principio de uniformidad, i.e. la idea de que todos los conceptos presentan una estructura similar (Weiskopf 2009a). Así, se desarrollaron teorías que postulan un único tipo de estructura conceptual. Hasta aproximadamente la década de 1970, resultó predominante la teoría clásica de las definiciones, según la cual los conceptos están constituidos por un conjunto de conceptos que expresan las condiciones necesarias y suficientes para caer bajo el concepto en cuestión. Esta teoría fue blanco de objeciones insuperables. Por un lado, se cuestionó la realidad psicológica de las definiciones, que resultan muy difíciles de formular y, probablemente, sólo sean posibles para un grupo muy limitado de conceptos, por ejemplo, los matemáticos. Por otro lado, se reportaron resultados de estudios de categorización que indican que las representaciones que tenemos de las categorías exhiben una estructura de rasgos estadísticos que recogen las tendencias centrales de los miembros de las categorías, sin que haya un conjunto de rasgos necesarios y

suficientes que todos los miembros compartan (Rosch 1975, Rosch & Mervis 1975).

A partir de estas críticas se fueron desarrollando enfoques con propuestas alternativas de estructura conceptual. Uno de ellos, el enfoque de prototipos, postula que los conceptos están estructurados como prototipos, que son estructuras de rasgos típicos o diagnósticos para pertenecer a una categoría (Hampton 1979, Rosch & Mervis 1975). Otra propuesta de estructura conceptual que surgió como alternativa a las definiciones fue el enfoque de ejemplares, según el cual los conceptos son conjuntos de representaciones de instancias de una categoría (Medin & Schaffer 1978, Nosofsky 1986, Nosofsky & Palmeri 1997). Una de las motivaciones principales de este enfoque es reflejar la influencia que tienen nuestros encuentros con instancias de categorías (nuestro conocimiento de instancias específicas) en nuestro sistema conceptual y en el uso que hacemos de los conceptos. Los enfoques de prototipos y ejemplares suelen considerarse emparentados, puesto que ambos sostienen que los conceptos refieren en virtud de una relación de similitud entre las representaciones conceptuales y las instancias que representan. Analizaré estos enfoques en el capítulo 2.

Un tercer enfoque respecto de la estructura conceptual, el de la llamada teoría-teoría de conceptos, se opone al énfasis que los enfoques de prototipos y ejemplares depositan en los rasgos superficiales de las categorías, así como a la idea de que los conceptos son meros conjuntos de rasgos correlacionados y propone, en cambio, que la estructura conceptual es análoga a la de las teorías científicas, incorporando relaciones causales y principios teóricos, así como representaciones de esencias ocultas (Ahn 1998, Ahn & Luhmann 2004, Carey 1991, Gopnik & Meltzoff 1997, Rehder & Burnett 2005). Me ocuparé de este enfoque en el capítulo 3. Por último, en oposición a todas las teorías que postulan una estructura interna para los conceptos (ya sea de rasgos definicionales, rasgos estadísticos, conjuntos de ejemplares o rasgos relacionados causalmente), el atomismo conceptual sostiene que la mayoría de los conceptos léxicos no tiene estructura interna, sino que son atómicos y su contenido queda determinado exhaustivamente por relaciones nomológicas mente-mundo, las cuales no involucran constitutivamente otros conceptos (Fodor 1999, 2008). Me ocuparé de este enfoque en el capítulo 4.

En el contexto de los debates en torno a la estructura informacional de los conceptos, se llevaron a cabo numerosos estudios empíricos que mostraron una serie de efectos, especialmente en tareas de categorización, que se usaron como evidencia a favor de uno u otro tipo de estructura conceptual. En particular, se buscaban efectos que fueran a la vez especialmente fáciles de explicar por medio de un tipo de estructura y difíciles de explicar por otros tipos de estructura. Estos estudios generaron un conjunto de evidencia de la cual

cualquier teoría de conceptos adecuada debería poder dar cuenta. Una consecuencia no sorprendente es que estos mismos datos sirvieron para debilitar a las teorías que adherían al principio de uniformidad, puesto que mostraron que ninguna de ellas era capaz de acomodar todos los efectos observados.

En consecuencia, muchos autores propusieron combinaciones de algunas de las estructuras, como teorías y prototipos (Keil 1989a, Keil, Smith, Simons & Levin 1998) o definiciones y prototipos (Osherson & Smith 1981), dando lugar a las primeras teorías híbridas. Actualmente, predominan los intentos por unificar distintos tipos de estructuras conceptuales en una misma teoría de conceptos, y el debate principal se da entre enfoques que defienden distintos modos de articular las diversas estructuras. Machery (2009, 2010, Machery & Seppälä 2011) reconstruye los debates actuales distinguiendo dos enfoques predominantes, uno híbrido y otro pluralista. Según esta distinción, las teorías híbridas proponen que los conceptos tienen *partes* constituidas por estructuras de tipos diferentes (Anderson & Betz 2001, Keil *et al.* 1998, Margolis & Laurence 2007, 2010, Osherson & Smith 1981), mientras que las teorías pluralistas proponen que cada categoría está representada por diversos conceptos independientes entre sí, cada uno de los cuales posee una estructura de tipo diferente (Machery 2009, Weiskopf 2009a). Así, usualmente se interpretan los enfoques híbrido y pluralista como dos enfoques en competencia. Más allá de las diferencias particulares entre las distintas teorías que abandonan el principio de uniformidad, se las agrupa adhiriendo a alguno de esos dos enfoques.

Como argumentaré en el capítulo 6, estos enfoques no están en competencia, sino que comparten sus tesis centrales. La diferencia principal entre ambos radica en el nivel de individuación de conceptos. Los enfoques híbridos enfatizan la relevancia de las asociaciones entre las distintas estructuras que representan una misma categoría, y, en este sentido, apelan a un nivel semántico de individuación que resulta en la denominación de todas las estructuras correferenciales como *partes* de un mismo concepto. Sostendré que el criterio para considerar distintas estructuras como partes de un mismo concepto es el hecho de que refieren a la misma categoría, en lugar de alguna relación estructural entre ellas. Los enfoques pluralistas, en cambio, enfatizan la posibilidad de activar cada una de las estructuras semánticas de modo independiente en la memoria de trabajo para representar una categoría y, por lo tanto, describen cada una de esas estructuras como un concepto independiente. En este sentido, considero que apelan a un nivel “arquitectural” de individuación, que se relaciona con la independencia estructural que mantienen las distintas representaciones entre sí. Al activar una de ellas para pensar acerca de una categoría, no necesariamente se activa el resto de las

representaciones correferenciales.

Sin embargo, sostendré que ambos enfoques coinciden, por un lado, en postular una multiplicidad de representaciones correferenciales que pueden usarse de modo independiente en los procesos cognitivos y, por otra parte, en incorporar condiciones de asociación entre las representaciones correferenciales, que están vinculadas en la memoria de largo plazo como representaciones de la misma categoría. Así, ambos enfoques comparten, en un sentido general, la tesis arquitectural según la cual las distintas estructuras correferenciales se pueden activar cada una independientemente del resto, y la tesis semántica según la cual podemos tener una multiplicidad de representaciones conceptuales correferenciales que difieren entre sí en el tipo de estructura interna. Uno de los desafíos que enfrentan las propuestas de este tipo, al que intentaré responder en esta tesis, consiste en articular, precisamente, las tesis semánticas y las tesis arquitecturales, resolviendo de algún modo la tensión entre los dos niveles de individuación de conceptos.

Hasta aquí he ofrecido un panorama de las posiciones en torno a cómo estaría estructurado el contenido de las representaciones mentales. Como mencioné, las representaciones mentales poseen un contenido (aquello acerca de lo cual son representaciones) y un vehículo. Los vehículos o formatos son los portadores del contenido. A diferencia del estado actual del debate en relación a las estructuras semánticas, en el cual predominan las propuestas híbridas y pluralistas, en relación al vehículo de los conceptos, en cambio, las principales perspectivas contemporáneas postulan un único tipo de vehículo.² Si bien no todas las teorías adoptan una postura explícita en este debate, pueden identificarse principalmente dos vertientes: aquella según la cual el formato de los conceptos es modal puesto que corresponde al formato de cada uno de los sistemas perceptivos (Barsalou 1999, Barsalou, Simmons, Barbey & Wilson, 2003, Prinz 2002), y aquella según la cual los conceptos están vehiculados por un formato amodal que no guarda ningún tipo de similitud estructural ni comparte recursos neuronales con los estados perceptivos. Según esta última el vehículo de los conceptos es análogo a los lenguajes naturales, pero sin ser ninguno de ellos (Fodor 1975, 1999, 2008, Fodor & Pylyshyn 1995). Tradicionalmente, las teorías de conceptos han adoptado la vertiente amodal, que propone una distinción entre los sistemas que sustentan la cognición y aquellos que sustentan la percepción. Sin embargo, en los últimos años se ha generado evidencia considerable a favor de que los conceptos pueden estar vehiculados por representaciones perceptivas y motoras (e.g. Pecher, Zeelenberg & Barsalou,

² Pueden mencionarse como excepciones a Dove (2009) y Carey (2009).

2003, 2004). Me ocuparé de esto en los capítulos 4 y 5.

En esta tesis me propongo defender una propuesta pluralista no sólo respecto de las estructuras semánticas, sino también de los vehículos conceptuales. Uno de los mayores desafíos para este enfoque pluralista es mostrar que puede heredar las ventajas de los distintos tipos de estructuras, sin heredar sus limitaciones. En este sentido, por ejemplo, es necesario establecer cómo se individualizan los conceptos y cómo se vinculan las distintas estructuras en los procesos mentales conceptuales, así como qué relación mantienen las representaciones de distintos formatos, e.g. si se pueden combinar o no. El objetivo de esta tesis es ofrecer una propuesta pluralista, tanto semántica como de vehículos, que pueda responder a estos desafíos.

Si bien fui adelantando acerca de qué tratará cada capítulo, a continuación presentaré el plan de la tesis de manera global y luego expondré brevemente lo que defenderé a lo largo de cada capítulo. En el capítulo 1 analizaré los requisitos explicativos que se han propuesto en la literatura sobre conceptos y seleccionaré entre ellos un conjunto de requisitos fundamentales que me permitirán, en los capítulos siguientes, evaluar las distintas teorías de conceptos. Dado que considero que una teoría adecuada de conceptos debe dar cuenta tanto de fenómenos relacionados con el poder expresivo del pensamiento como de fenómenos relacionados con el uso de conceptos en procesos cognitivos, propondré seleccionar como requisitos fundamentales: el contenido, el alcance, la composicionalidad, la categorización, la adquisición y la publicidad.

En los capítulos 2 a 4 me ocuparé de reconstruir y evaluar las principales teorías de conceptos que adhieren a la tesis de la uniformidad. Estos capítulos tienen por objetivo establecer las ventajas y desventajas de cada una de las estructuras conceptuales propuestas en la literatura, en relación a la explicación de los distintos requisitos. Un modo tradicional de reconstruir los debates en el campo de los conceptos es en torno a grandes perspectivas teóricas: la teoría clásica de las definiciones, las teorías de prototipos, de ejemplares, de teorías y el atomismo. Como señala Murphy (2002) este tipo de reconstrucción puede resultar engañosa en, por lo menos, tres sentidos. Primero, puede generar la falsa impresión de que no hubo progresos en este campo y que los debates entre las distintas teorías siguen siendo los mismos que se sostenían en el momento del surgimiento de los principales enfoques. Segundo, ciertos fenómenos no parecen tener lugar en el marco de determinadas teorías, e.g. la categorización en el atomismo, o la combinación de conceptos en la perspectiva de ejemplares. Por último, la investigación disponible sugiere que todas las perspectivas en disputa tienen dificultades, todas ellas son incorrectas en mayor o menor medida. Como

señala Murphy, para dar cuenta de todos los efectos observados, probablemente se necesita un enfoque más inclusivo.

Si bien los primeros capítulos de la tesis tendrán esta estructura criticada por Murphy, la intención no es reconstruir los debates entre las teorías, sino destacar los fenómenos a los que cada una se ajusta mejor, para justificar así su inclusión en una teoría pluralista como la que me propongo defender. La organización en torno a enfoques teóricos es simplemente un modo de presentar los estudios disponibles, que constituyen el conjunto de evidencia que se debe tener en cuenta para formular una teoría de conceptos y que, en su gran mayoría, fueron llevados a cabo como parte de la defensa de alguna de las perspectivas principales mencionadas. Así, el objetivo no será defender una u otra perspectiva, sino evaluar en cada caso la evidencia y los argumentos a su favor, así como sus limitaciones. Me interesa mostrar, por un lado, qué es lo que motiva su inclusión en una propuesta pluralista y, por otro lado, qué dificultades acarrea cada una, para las cuales una propuesta pluralista debe ofrecer soluciones. El orden en el que me ocuparé de cada uno de los enfoques principales respetará a grandes rasgos el orden temporal en el que surgieron. De esta manera, puede verse a qué aspectos de los enfoques previos se opone cada uno de ellos, o a qué tesis particulares pretenden ofrecer alternativas.

En el capítulo 2 me ocuparé de los primeros enfoques que atribuyen a los conceptos una estructura de información interna: el de las definiciones, el de prototipos y el de ejemplares. Con respecto a la teoría clásica de las definiciones, que sostiene que los conceptos están estructurados como conjuntos de condiciones necesarias y suficientes para pertenecer a una categoría, señalaré que su principal desventaja es que su alcance es muy limitado. Si bien resulta adecuada para algunos conceptos, distintos estudios muestran que, para la mayoría de los conceptos, no disponemos de condiciones necesarias y suficientes. Respecto al enfoque de prototipos, según el cual los conceptos son estructuras de rasgos típicos o diagnósticos para pertenecer a una categoría, señalaré que resulta especialmente adecuado para explicar los efectos de tipicidad observados en las tareas de categorización, pero no otros efectos reportados y, si bien este enfoque ofrece explicaciones adecuadas de la adquisición, tiene dificultades especialmente en relación al requisito de composicionalidad.

El enfoque de ejemplares, según el cual los conceptos son conjuntos de representaciones de instancias particulares de una categoría, se encuentra en una situación similar al de prototipos, dado que resulta adecuado para explicar ciertos fenómenos observados en las tareas de categorización (pero no todos), y para explicar la adquisición. Sin embargo, enfrenta dificultades principalmente en relación al alcance, la publicidad y la

composicionalidad. En el capítulo 3 abordaré las distintas propuestas que se agrupan en el enfoque denominado de teoría-teoría, según el cual los conceptos están estructurados como conjuntos de rasgos relacionados causalmente y pueden incorporar representaciones de esencias ocultas, así como principios teóricos dependientes de cada dominio de representación (tales como el de los objetos físicos, los agentes, o los seres vivos). Sostendré que este enfoque ofrece ventajas en la explicación de algunos efectos observados en la categorización, el alcance y la adquisición de conceptos. Sin embargo, tiene limitaciones para explicar el contenido de los conceptos y la publicidad.

En el capítulo 4 reconstruiré las tesis principales del atomismo conceptual fodoriano, según el cual la mayoría de los conceptos léxicos son atómicos y su contenido queda exhaustivamente determinado por relaciones nomológicas mente-mundo que no involucran constitutivamente otros conceptos, y evaluaré la satisfacción de cada uno de los requisitos fundamentales. Sostendré que si bien el atomismo resulta especialmente adecuado para dar cuenta de los requisitos de alcance, composicionalidad y publicidad, tiene sin embargo dificultades con el requisito de intencionalidad a la hora de dar cuenta de la individuación de conceptos coextensivos. Por otra parte, sostendré que el atomismo no ofrece una explicación adecuada de la adquisición, puesto que para dar cuenta de ella sería necesario explicitar el vínculo que se plantea entre los conceptos atómicos y los mecanismos de acceso semántico que son intermediarios entre los conceptos y el mundo. Un problema similar se presenta respecto de la categorización. Si bien se podría articular la metafísica fodoriana de los conceptos con explicaciones de la categorización, para esto sería necesario esclarecer la relación entre los conceptos atómicos y otros tipos de estructuras de información, tales como los prototipos, cuya mediación es necesaria para la adquisición y presumiblemente también para la categorización. Como mencioné, no todas las teorías se comprometen con tesis específicas respecto del vehículo de los conceptos. La teoría fodoriana constituye uno de los desarrollos más detallados relativos a los vehículos amodales, de modo que en este capítulo me ocuparé también de evaluar las ventajas y limitaciones de este tipo de vehículo.

En el capítulo 5 abordaré el neo-empirismo de conceptos. Reconstruiré críticamente la teoría de proxitipos de Prinz (2002), la cual abandona parcialmente la tesis de la uniformidad, en tanto afirma que los conceptos pueden tener distintos tipos de estructuras de información (pueden, por ejemplo, incluir rasgos prototípicos o conjuntos de ejemplares), pero mantiene esta tesis con respecto a los vehículos, dado que afirma que todos los conceptos están vehiculados por representaciones perceptivas. En este capítulo destacaré las ventajas de la incorporación de distintos tipos de estructuras de información, en particular

para dar cuenta del requisito de categorización, así como las ventajas que implica la adopción de representaciones perceptivas para explicar la adquisición de conceptos. Sin embargo, mostraré que no consigue dar cuenta de la individuación del contenido de los conceptos, ni de la representación de conceptos teóricos y abstractos recurriendo únicamente a representaciones perceptivas.

De este modo, el análisis en los capítulos 2 a 5 de las distintas teorías que adhieren, aunque sea parcialmente, a la tesis de la uniformidad pretende, por un lado, fundamentar el abandono de esta tesis tanto respecto de las estructuras de información como respecto de los vehículos conceptuales y, por otro lado, fundamentar qué tipos de estructuras semánticas y de vehículos debería incorporar una propuesta de tipo pluralista. Para establecer las tesis centrales de una propuesta pluralista, en el capítulo 6 me ocuparé de modo general, de las propuestas híbridas, pluralistas y eliminativistas. En primer lugar, discutiré la distinción que establecen Machery (2009) y Weiskopf (2009a) entre teorías híbridas y pluralistas. Mostraré, por un lado, que la distinción que proponen les atribuye a las teorías híbridas compromisos que estas teorías no adoptan y, por otro lado, que las llamadas teorías híbridas y las llamadas pluralistas comparten ciertas tesis centrales.

En segundo lugar, sugeriré que la adopción de esta distinción, fundamentalmente por parte de los defensores de las teorías híbridas, puede deberse a un intento por resistir la asociación que propone Machery entre pluralismo y eliminativismo. Machery (2009) sostiene que hay evidencia a favor de la existencia de prototipos, ejemplares y teorías, y, en este sentido, establece una forma de pluralismo, pero argumenta que estas entidades no comparten propiedades relevantes, por lo que no constituyen una clase natural y, por lo tanto, el término “concepto” debe ser eliminado del vocabulario científico. Reconstruiré los argumentos de Machery a favor del eliminativismo y sugeriré una respuesta posible a estos argumentos que muestra que el eliminativismo no se sigue del pluralismo. De este modo, sugeriré cuáles son las tesis centrales de una propuesta pluralista que no implica la eliminación de la noción de “concepto”.

En el capítulo 7 desarrollaré una propuesta que postula una pluralidad de representaciones mentales de distinto tipo, a la que denomino *pluralismo amplio*, puesto que abarca tanto un pluralismo de estructuras semánticas como un pluralismo de vehículos. Este enfoque pluralista presenta numerosas ventajas frente a las posturas teóricas que adhieren a la tesis de la uniformidad, puesto que hereda las ventajas explicativas de las distintas teorías. Mostraré que puede enfrentar las dificultades que presenta cada teoría por separado y detallaré cómo puede dar cuenta de cada uno de los requisitos fundamentales. Asimismo,

sugeriré qué tipo de arquitectura cognitiva resulta adecuada para explicar la articulación entre los distintos tipos de vehículos conceptuales. Sugeriré que el pluralismo amplio se puede combinar con una arquitectura de procesamiento dual. Así, intentaré dar respuestas a los principales desafíos que enfrentan las propuestas pluralistas.

Capítulo 1

Requisitos explicativos de una teoría de conceptos

Es habitual que las teorías de conceptos se organicen en torno a *desiderata* o requisitos explicativos, esto es, que partan de la enumeración de una serie de fenómenos de los cuales la teoría pretende dar cuenta. Sin embargo, no hay un acuerdo total entre las distintas teorías respecto de estos requisitos. Aunque hay propuestas teóricas que consideran la mayoría de los fenómenos propuestos en la literatura, e.g. la teoría de proxitipos de Prinz (2002), o la propuesta pluralista de Weiskopf (2009a), muchas teorías se ocupan sólo de algunos de ellos. Los fenómenos de los que se ocupan las teorías de conceptos pueden presentarse en dos grandes grupos. Por un lado, se pueden identificar fenómenos relacionados con el poder expresivo del pensamiento: el contenido, el alcance y la composicionalidad. Por otro lado, se pueden identificar fenómenos relacionados con la adquisición y el uso de los conceptos, i.e. con los procesos cognitivos en los que intervienen los conceptos, fundamentalmente la adquisición, el cambio conceptual, la categorización, el razonamiento y la publicidad. En este capítulo desarrollaré cada uno de los requisitos considerados usualmente en la literatura, con el objetivo de seleccionar un conjunto de requisitos explicativos que una teoría de conceptos debería satisfacer. En §1 me ocuparé de los requisitos relacionados con el poder expresivo del pensamiento y en §2 de aquellos relacionados con los procesos cognitivos conceptuales. Por último, en §3, señalaré qué requisitos considero fundamentales para una teoría adecuada de conceptos. La adopción de este conjunto de requisitos fundamentales me permitirá, en los capítulos siguientes, evaluar las distintas teorías de conceptos.

1. Requisitos relacionados con el poder expresivo del pensamiento

Las teorías de conceptos coinciden en afirmar que los conceptos son los constituyentes de los pensamientos (Margolis & Laurence 2014). Como consecuencia de esto, una teoría de conceptos debería dar cuenta de los fenómenos relacionados con el poder expresivo del pensamiento: contenido, alcance y composicionalidad. En lo que sigue desarrollaré cada uno de estos requisitos.

Intencionalidad

Una teoría de conceptos debe explicar en qué consiste que un concepto sea acerca de algo, i.e. tenga contenido intencional. Para ello, es necesario establecer qué tipo de

información portan los conceptos, por ejemplo, si es conocimiento proposicional, estadístico, o de otro tipo y si la información está estructurada o es atómica, y qué tipo de relación tienen con el mundo, por ejemplo, si se trata de una relación causal o de similitud, así como qué relación tienen con los otros conceptos. Los enfoques que analizaré en los próximos capítulos coinciden en adoptar, aunque sea de modo implícito, alguna versión de semántica externalista, i.e. asumen que la individuación del contenido depende, por lo menos en parte, de relaciones con el entorno, en lugar de postular que el contenido está completamente determinado por propiedades intrínsecas de los individuos. Así, asumen que el contenido de los conceptos está determinado por los conjuntos de entidades a los que refieren.

Las teorías semánticas tradicionales más desarrolladas respecto del contenido de los conceptos son la semántica informacional (Dretske 1981, Fodor 1990) y la semántica de rol conceptual (Block 1986). Según la semántica informacional, el contenido consiste en información acerca del mundo (propiedades o eventos) y los conceptos portan esa información en virtud de relaciones de covariación nomológica con sus referentes. Según la semántica de rol conceptual, en cambio, el contenido está determinado por los roles inferenciales de los conceptos, i.e. por las inferencias en las que estos participan. Así, una diferencia fundamental entre ellas radica en que según la semántica informacional la individuación del contenido de un concepto es independiente del resto de los conceptos que posee un individuo, mientras que según la semántica de rol conceptual el contenido de un concepto se determina en parte por las relaciones que mantiene con otros conceptos.

El contenido de los conceptos resulta central para su individuación. Los conceptos son acerca de algo y en general se los identifica, por lo menos en parte, por aquello que representan. Pues bien, distintos autores proponen que el contenido intencional no es suficiente para la individuación de conceptos y que es necesario incorporar un requisito sobre el contenido cognitivo (Smith & Minda 1998, Prinz 2002, Putnam 1975, Weiskopf 2009b). Este tipo de contenido permitiría dar cuenta de los llamados casos Frege, i.e. casos de distintos conceptos con idéntico contenido intencional, y los casos de tierra gemela, en los cuales un mismo concepto tiene distintos contenidos intencionales en distintos mundos posibles. Este tipo de contenido se relaciona generalmente con las concepciones epistémicas de los individuos y se propone para permitir una individuación de conceptos de grano más fino que la que permitiría únicamente el contenido intencional. Así se explicaría, por un lado, cómo en los casos Frege un individuo podría tener creencias contradictorias que involucren conceptos que refieren a la misma propiedad en el mundo (e.g. Luisa Lane cree que Superman vuela, y a la vez cree que Clark Kent no vuela) y, por otro lado, cómo en los casos

de tierra gemela dos individuos pueden tener creencias y comportamientos idénticos en relación a propiedades distintas (e.g. en relación a H₂O en la tierra actual y a XYZ en la tierra gemela).

Algunos autores, como Fodor (1987), propusieron que los casos de tierra gemela se explican postulando un contenido estrecho (que, a diferencia del cognitivo, no es epistémico), el cual se define en función del contenido intencional o amplio. Mientras que el contenido amplio está constituido por propiedades en el mundo y determinado por relaciones nomológicas, el contenido estrecho es una función de contextos a condiciones de verdad o a contenido amplio (Skidelsky, en prensa).³ Así, por ejemplo, AGUA en el contexto de esta tierra se proyecta en H₂O, mientras que en el contexto de la tierra gemela se proyecta en XYZ. Un individuo en esta tierra y su gemelo, pueden tener pensamientos acerca de agua con el mismo contenido estrecho, dado que, si estuvieran en el mismo contexto, sus pensamientos tendrían las mismas condiciones de verdad.

Otros filósofos, como Block (1986), consideran que una semántica de rol conceptual puede dar cuenta del contenido cognitivo. En versiones más actuales, en cambio, el contenido cognitivo se explica apelando a la misma teoría semántica que explica el contenido intencional, y ambos se determinan por relaciones con propiedades en el mundo (Prinz 2002, Weiskopf 2009b). También se han propuesto versiones mixtas, que postulan una semántica informacional para la determinación del contenido intencional y una semántica de rol conceptual para la determinación del contenido cognitivo (Carey 2009).

Dado que el contenido intencional no resulta suficiente para la individuación de conceptos, consideraré el requisito de contenido cognitivo como parte del requisito de contenido intencional. Sin embargo, también señalaré que postular dos tipos de contenido requiere especificar cómo se articulan ambos en la individuación de conceptos, cualquiera sea la teoría semántica que se postule para cada tipo de contenido. La individuación de conceptos por una teoría semántica de doble factor puede resultar problemática, como analizaré en el capítulo 5, a partir de la teoría de proxitipos.

Alcance

Somos capaces de pensar acerca de perros, manzanas y tractores, pero también acerca

³ Fodor (1994) abandona la noción de contenido estrecho y propone tratar los casos de tierra gemela como casos nomológicamente imposibles. Con respecto a los casos de conceptos coextensivos, propone que estos se individúan por sus vehículos. En el capítulo 4 presentaré una serie de objeciones a esta propuesta.

de números, repúblicas, el miedo y la felicidad. Podemos representar propiedades de distinto tipo, tanto concretas como abstractas. Tenemos una variedad de conceptos que incluye, *inter alia*, conceptos de clases naturales, como AGUA, conceptos de artefactos, como TELEVISOR, conceptos derivados teóricamente, como PROTÓN y conceptos acerca de estados internos, como DOLOR (Prinz 2002). Al postular un requisito de alcance, se tiene en mente que para poder explicar nuestra capacidad representacional, una teoría de conceptos debe dar cuenta del amplio espectro de conceptos con los que nuestra mente puede operar.

En ocasiones, este requisito se superpone parcialmente con el requisito de intencionalidad, puesto que para satisfacer el requisito de alcance es preciso también poder dar cuenta de la determinación del contenido de los conceptos. Sin embargo, se trata de requisitos distintos. El requisito de alcance incluye consideraciones tanto respecto del contenido como de los vehículos de los conceptos y permite diferenciar las distintas teorías en función del espectro de conceptos a los que se aplican.

Composicionalidad

Nuestro pensamiento exhibe, al menos, dos fenómenos que una teoría de conceptos debe explicar. Por un lado, es productivo, esto es, podemos, a partir de un sistema representacional finito, generar una infinita cantidad de pensamientos nuevos. Por otro lado, es sistemático, esto es, si somos capaces de generar ciertos pensamientos, también somos capaces de generar otros relacionados, recombinando sus constituyentes. Margolis y Laurence (2007) recurren a un ejemplo de George Miller para ilustrar la productividad del pensamiento por analogía con la productividad del lenguaje. Si tomamos oraciones de 20 palabras y consideramos (por estipulación) que al construir una oración, cada palabra es elegida entre un conjunto de 10 palabras, esto permite calcular que somos capaces de generar y entender, por lo menos, 10^{20} oraciones de 20 palabras. Para otorgarle dimensión y perspectiva a este número, Margolis y Laurence señalan que éste equivale aproximadamente a mil millones de veces la cantidad de neuronas del cerebro y a mil veces la cantidad de segundos en la historia del universo. Si suponemos que, aproximadamente, cada oración se corresponde con un pensamiento diferente, la capacidad productiva del pensamiento resulta notable. Este es un fenómeno que requiere explicación en cualquier teoría de la mente y en particular del pensamiento. La sistematicidad también se puede ilustrar por analogía con los lenguajes naturales. Así como ningún hablante del castellano puede entender o producir la oración “El hombre cuida al perro” a no ser que también pueda entender o producir la oración “El perro cuida al hombre”, tampoco es plausible que una persona pueda pensar que

el hombre cuida al perro, pero que no pueda pensar que el perro cuida al hombre.

Pues bien, ¿cómo se relacionan estos fenómenos con las teorías de conceptos? Si se asume que los conceptos son los constituyentes de los pensamientos, se puede apelar a la composicionalidad como una explicación de estos dos fenómenos. Fodor define informalmente la composicionalidad afirmando que "... la sintaxis y el contenido de un concepto complejo se determinan normalmente por la sintaxis y el contenido de sus constituyentes" (1999, p. 135). Respecto de la productividad, la representación de una cantidad infinita de conceptos por una mente con un sistema representacional finito se logra si cada concepto complejo se puede descomponer finitamente en la sintaxis y los contenidos de sus constituyentes. El argumento para explicar la sistematicidad es similar. El hecho de que una mente que puede tener el pensamiento de que el hombre cuida al perro, también puede tener el pensamiento de que el perro cuida al hombre se explica por medio de la composicionalidad porque el contenido de los pensamientos complejos queda determinado de manera composicional por el contenido de sus constituyentes y las reglas de combinación sintáctica. De este modo, para poder explicar la productividad y la sistematicidad, los conceptos deben satisfacer el requisito de composicionalidad.

Otro modo de presentar este requisito es en términos del *Requisito de Generalidad* de Evans (1982), según el cual los conceptos se deben poder recombinar libremente:

Si a un sujeto se le puede atribuir el pensamiento de que a es F , entonces debe tener los recursos conceptuales para tener el pensamiento de que a es G , para cualquier propiedad de ser G de la cual tiene una concepción. (1982, p. 104)

Evans presenta este requisito en términos de habilidades conceptuales, dado que sostiene que los conceptos son habilidades, en lugar de representaciones mentales. Sin embargo, arriba al requisito de generalidad a través de consideraciones análogas a las que proponen Fodor y Pylyshyn (1995) sobre las relaciones sistemáticas entre pensamientos.

En general, hay acuerdo entre los autores en que la explicación de los fenómenos de productividad y sistematicidad requiere un sistema de conceptos que se puedan combinar para formar conceptos más complejos. Sin embargo, el requisito de composicionalidad se ha interpretado de distintas maneras. La propuesta de Fodor es la que se denomina versión fuerte de este requisito, en tanto propone que el contenido de los conceptos complejos se agota en los contenidos de los conceptos constituyentes y las reglas de combinación y, a la vez, que los conceptos complejos heredan todo el contenido de los conceptos constituyentes. Así, el

contenido de los conceptos simples es aquello que contribuyen a los conceptos complejos (Fodor 1999). Corresponde aclarar que el requisito de composicionalidad no suele entenderse de manera universal como sosteniendo que los conceptos se componen *siempre*. Ni siquiera Fodor y Lepore (1996), tenaces defensores de este requisito, niegan los casos de modismos o expresiones idiomáticas. “Caer en la cuenta” o “al pie de la letra” son expresiones cuyo contenido no depende de la composicionalidad semántica ni de la estructura sintáctica. El contenido de este tipo de expresiones debe adquirirse como un bloque, i.e. como si correspondiera a un concepto léxico. En ese sentido, estos no cuentan como casos de combinación de conceptos, sino que funcionarían más bien como un solo concepto, sin una estructura de constituyentes semánticos. Sin embargo, éste no puede ser el caso general, puesto que de esa manera o bien nuestro sistema representacional debería ser infinito, o bien no tendríamos la capacidad para generar infinitos pensamientos nuevos. Puesto que no se da ninguna de las dos cosas, las expresiones idiomáticas no pueden ser el caso general, sino la excepción.

Más allá de las expresiones idiomáticas, respecto de las cuales hay acuerdo en que su contenido no se determina de modo composicional, algunos autores proponen entender al requisito de composicionalidad en un sentido más débil al que defiende Fodor. Prinz (2002) y Robbins (2002), entre otros, proponen una lectura debilitada, según la cual para explicar la productividad y la sistematicidad, debe ser *posible* combinar conceptos de modo composicional, pero puede haber otras formas de combinación que se usen más frecuentemente, en las cuales la interpretación semántica de los constituyentes no agote el significado de la representación compleja. Así, la lectura debilitada es compatible con modelos de combinación de conceptos, como el modelo de Costello y Keane (2000), que explican el fenómeno de las propiedades emergentes. Este fenómeno se observa en algunos casos en los que, al combinar conceptos, el resultado tiene propiedades que no se atribuyen a ninguno de los conceptos iniciales. Un ejemplo de este fenómeno muy discutido en la literatura es el de los CARPINTEROS EGRESADOS DE HARVARD, a quienes se atribuye el no ser materialistas, a pesar de que aquel no es un rasgo de CARPINTERO ni de EGRESADO DE HARVARD. En este caso, el no ser materialista es una propiedad emergente (Kunda, Miller & Claire 1990).

Es importante notar que tanto en la lectura fuerte como en la débil, el requisito de composicionalidad se propone como la mejor explicación para los fenómenos de productividad y sistematicidad. Un sistema finito de representaciones conceptuales debe ser suficiente para dar cuenta de la infinita cantidad de pensamientos que podemos tener y la

mejor manera de explicar este fenómeno es apelando a la combinación de conceptos. Lo que distingue la lectura fuerte de la débil es el alcance del requisito. En la lectura fuerte se sostiene que para todos los casos de combinación de conceptos el contenido de los conceptos complejos queda exhaustivamente determinado por el contenido de los conceptos constituyentes y las reglas de combinación, al mismo tiempo que aquello que un concepto simple contribuye a un concepto complejo agota su contenido. En la lectura débil, en cambio, se restringe la composicionalidad a casos límite, i.e. la composicionalidad garantiza la productividad y la sistematicidad, pero se incorporan otras formas de combinación no composicionales. Así, ya sea en su interpretación fuerte o en la versión más débil, la composicionalidad parece ser un requisito central para las teorías de conceptos.

2. Requisitos relacionados con el uso de los conceptos en procesos cognitivos

Uno de los objetivos explicativos de la mayoría de las teorías de conceptos es dar cuenta de cómo adquirimos conceptos y cómo contribuyen los conceptos en una serie de capacidades cognitivas.⁴ En este sentido, una teoría de conceptos debería poder dar cuenta, por un lado, de los requisitos de adquisición y cambio conceptual, y, por otro lado, de los de categorización, razonamiento y publicidad. A continuación desarrollaré estos requisitos.

Adquisición

Según este requisito, una teoría de conceptos debe incluir especificaciones acerca de cómo estos se adquieren. Debe explicar si los conceptos son innatos o aprendidos y cuáles son los mecanismos de adquisición o aprendizaje. El debate entre las teorías empiristas e innatistas se da en torno de esta cuestión. Las primeras, por ejemplo la teoría de proxitipos de Prinz (2002), proponen que la gran mayoría de los conceptos son aprendidos y postulan para esto mecanismos de dominio general. En cambio, las teorías innatistas, como la de Fodor (1975), postulan una base de conceptos innatos y recurren para explicar la adquisición a mecanismos de dominio específico.⁵

Las teorías empiristas proponen mecanismos de adquisición que, en general, son epistémicos, en tanto se basan en una relación racional con el mundo. Se supone que estos mecanismos son análogos a la formulación y testeo de hipótesis científicas, de modo que se

⁴ Se podría mencionar a la teoría de conceptos de Fodor (1999) como una excepción, dado que excluye los fenómenos psicológicos relativos al uso de los conceptos de aquello que la teoría pretende explicar. Diré algo más sobre esto en §3 y me ocuparé de esta teoría en el capítulo 4.

⁵ Aunque en Fodor (1999) ya no se habla de conceptos innatos sino de mecanismos conceptuales innatos. Volveré sobre esto en el capítulo 5.

establece una relación de evidencia entre la experiencia y los conceptos. La adquisición de conceptos podría basarse así en mecanismos inductivos de dominio general (Prinz, 2002). Por ejemplo, los procesos de adquisición de prototipos, por medio del registro de regularidades estadísticas (como la coocurrencia de propiedades en el mundo), involucran procesos de este tipo. Según las teorías innatistas, en cambio, la adquisición no es una forma de aprendizaje, porque no involucra una relación racional ni evidencial con la experiencia. Se trata, en cambio, de un proceso causal bruto, del cual no se puede dar cuenta en términos psicológicos, de disparo de conceptos ante los estímulos correspondientes en el mundo (Fodor 1999, Fodor 2008).

Se han propuesto también versiones que combinan en cierta medida aspectos innatistas y empiristas en la adquisición de conceptos. Es el caso de la propuesta de Carey (2009), que propone que tenemos un conjunto de representaciones conceptuales innatas (en tanto se producen como *outputs* de sistemas innatos ante encuentros con los estímulos relevantes), a las que denomina “cognición nuclear”. A partir de estas representaciones, nuestras interacciones con el entorno y, en algunos casos, con intervención del lenguaje natural, se adquiere el resto de los conceptos por procesos racionales. Retomaré esta propuesta en el capítulo 7.

En muchos casos el requisito de adquisición se aborda en relación al fenómeno del cambio conceptual, que describiré a continuación.

Cambio conceptual

Una de las preguntas centrales que guía la investigación empírica en torno a los conceptos es si los conceptos de los niños y de los adultos son similares en contenido y en estructura o si difieren cualitativamente unos de otros. Para responder esta pregunta, se han elaborado, por un lado, estudios que evalúan las diferencias en los juicios de categorización en distintas edades y, por otro lado, estudios que atienden al desarrollo conceptual en los primeros años de vida. Mandler (2004) señala ciertos fenómenos que requieren explicación:

[...] por qué los bebés de tres meses pueden ver la diferencia entre un perro y un gato, mientras que no los diferencian en algunos tests hasta meses después, por qué los bebés pasan ciertos tests numéricos en los que fallan en edad preescolar, o por qué parecen tener conocimiento sobre física que aparentemente abandonan unos años después. (p. viii)

Mandler propone explicar estos fenómenos a partir de la continuidad entre los conceptos de los niños y los de los adultos. Señala que las diferencias observadas a lo largo del desarrollo se deben a un enriquecimiento y diferenciación a partir de representaciones más bien abstractas o globales, y no a una reorganización fundamental del sistema conceptual. Adoptando un enfoque análogo, Murphy (2002) señala que las representaciones conceptuales y los procesos de adquisición de conceptos son similares en niños y adultos, y que las diferencias en el desempeño en ciertas tareas (particularmente en la categorización) pueden explicarse apelando a diferencias en sus experiencias con miembros de las categorías, en el conocimiento de dominios específicos o en la capacidad de procesamiento.

Carey (2009), en cambio, propone que, para algunos dominios, se produce una discontinuidad conceptual entre los sistemas de los niños y los adultos. Carey afirma que los sistemas de representación resultantes pueden ser inconmensurables con aquellos a partir de los cuales fueron construidos, en tanto los conceptos de uno no pueden ser expresados por los conceptos del otro. Por ejemplo, el concepto de los niños de FÍSICAMENTE REAL difiere del concepto adulto de MATERIAL, en tanto, los niños, e.g. atribuyen peso sólo a objetos de gran tamaño, mientras que afirman que los objetos muy pequeños pesan 0g y consideran que las sombras existen en completa oscuridad pero no las vemos.

Categorización

Generalmente, en la literatura sobre el tema, se caracteriza a la categorización como la capacidad que tenemos de identificar a qué categoría pertenece un objeto. Se diferencia de la representación en sí, puesto que se relaciona con la aplicación de un concepto a una cosa, y no meramente con nuestra capacidad de representar esa cosa. Prinz (2002, p. 9) señala que la categorización es la “contraparte epistémica” de la relación semántica de representación. Así, la categorización es una relación que se establece entre una categoría y un elemento del mundo cuando se *juzga* que dicho elemento cae bajo la categoría.⁶ Algunos autores, como Prinz (2002) incluyen también en este requisito la capacidad de anticipar las características que tendrá un objeto por pertenecer a una categoría. Otros autores, como Murphy (2002) o Machery (2009), consideran que la capacidad de anticipar características debe explicarse como parte de los procesos relacionados con las inferencias inductivas y la incluyen, por lo

⁶ Tanto en la literatura filosófica como psicológica el uso del término “categoría” es ambiguo. Algunos autores consideran que “categoría” es sinónimo de “concepto” (Fodor 1999). Otros, en cambio, establecen una diferencia señalando que las categorías son aquello que los conceptos representan (Murphy 2002, Prinz 2002, Weiskopf 2009a). En esta tesis utilizaré el término de manera indistinta, dado que en general el contexto aclara cuál es la interpretación adecuada.

tanto, dentro de este último requisito.

Los estudios psicológicos de tareas de categorización reportan diversos fenómenos que es preciso tener en cuenta al tratar este requisito. En especial, hay tres tipos de efectos cuyo estudio ha guiado gran parte de los desarrollos de los enfoques de prototipos, de ejemplares y la teoría-teoría. Se trata de los “efectos de tipicidad”, los “efectos de ejemplares” y los “efectos causales” (Murphy 2002, Machery 2009). Los efectos de tipicidad involucran una serie de asimetrías observadas en la categorización de distintos elementos de una categoría. Por un lado se observó en distintos estudios que los individuos juzgan algunos elementos como más típicos de una categoría que otros. Por ejemplo, para la categoría VEHÍCULO, los individuos consideran que un auto es un mejor ejemplo que una silla de ruedas y ésta a su vez es un mejor ejemplo que un ascensor (Rosch 1975). Por otra parte, Rosch y Mervis (1975) observaron una correlación directa entre la tipicidad de un ítem respecto a una categoría, y la cantidad de atributos que comparte con otros miembros de la categoría. En otros estudios, se observó una correlación directa entre el grado de tipicidad y la velocidad de categorización: generalmente las instancias más típicas de una categoría se categorizan más rápido que las menos típicas (Hampton 1997).

Otros estudios, sin embargo, señalan un efecto inverso, i.e. que instancias atípicas de una categoría, con las que los individuos están familiarizados, se categorizan más rápido que otras típicas (Medin & Schaffer 1978). También se ha observado un efecto relacionado en estudios de aprendizaje de categorías artificiales. Estos estudios generalmente incluyen una primera etapa de aprendizaje de la categoría a partir de algunos de sus elementos y una segunda etapa en la que se deben categorizar tanto elementos nuevos como elementos ya conocidos de la primera etapa. Los resultados de estos estudios sugieren que los individuos categorizan más rápido los elementos ya conocidos que los elementos nuevos, a pesar de que sean igualmente típicos de la categoría en función de los atributos que comparten (Smith & Minda 1998). Estos últimos efectos suelen llamarse “efectos de ejemplares” (Murphy 2002, Machery 2009), dado que son aquellos que las teorías de ejemplares predicen. Me ocuparé en detalle de estos efectos en el capítulo 2, en relación al análisis de los enfoques de prototipos y de ejemplares.

Asimismo, también se observaron una serie de efectos causales que sugieren que ciertos rasgos son más relevantes que otros a la hora de categorizar un elemento, en función de las relaciones de dependencia que mantienen con otros rasgos. En especial, aquellos rasgos que causan la ocurrencia de otros rasgos son más relevantes para la categorización que aquellos rasgos que son efectos (Ahn & Luhmann 2004, Rehder 2003). Por ejemplo, en uno

de los estudios en el que los participantes debían agrupar a Jane, que está deprimida porque tiene baja autoestima, con Susan, que está deprimida porque ha estado bebiendo, o con Bárbara, quien está a la defensiva porque tiene baja autoestima, la mayoría (73.6% de los participantes), eligió agrupar a Jane con Bárbara (Ahn & Luhman 2004). Esto sugiere que la similitud de causa (en este caso la baja autoestima) resultó más relevante que la similitud de efecto (en este caso la depresión). Me ocuparé de este tipo de efectos en el capítulo 3, al analizar el enfoque de teoría-teoría.

Se han realizado numerosos estudios en torno a estos tres grupos de efectos. Al postular el requisito de categorización como parte de una teoría de conceptos, se pretende explicar este proceso de un modo que permita dar cuenta de los distintos fenómenos empíricos observados.

Razonamiento: inferencias deductivas e inductivas

Tenemos la capacidad para, por ejemplo, inferir a partir de que puedo llegar en subte o colectivo y que no funciona el subte que, por lo tanto, debo tomar el colectivo. Una teoría de conceptos debería dar cuenta de la coherencia inferencial en ese tipo de razonamientos. Fodor y Pylyshyn (1995) señalan que lo que permite realizar inferencias deductivas de este tipo es la sensibilidad de estos procesos a la sistematicidad de nuestros pensamientos. Así, no es posible que una mente pueda inferir, por ejemplo, P de P&Q, pero no pueda inferir R de P&Q&R. Fodor y Pylyshyn no incluyen al razonamiento deductivo como un requisito explicativo para las teorías de conceptos, pero analizan las condiciones que la arquitectura cognitiva debe satisfacer para poder dar cuenta de este tipo de razonamiento y apelan a la sistematicidad de las inferencias para sostener que los conceptos deben constituir un sistema simbólico como el lenguaje del pensamiento. Si, como señalan Fodor y Pylyshyn, la capacidad para realizar inferencias deductivas se explica, en parte, por la sistematicidad del pensamiento y la mejor explicación para la sistematicidad es la composicionalidad, entonces para poder dar cuenta de las inferencias deductivas una teoría de conceptos debe satisfacer el requisito de composicionalidad.

Si bien no es necesario que una teoría de conceptos incluya una explicación detallada de los mecanismos inferenciales, los conceptos que postule deben permitir la coherencia inferencial. En términos de Weinberg (2003), deben ser “amigables al pensamiento”, en el sentido de poder instanciar, *inter alia*, inferencias deductivas. Así, este requisito impone ciertas restricciones respecto del vehículo de los pensamientos, en tanto exige que estos puedan tener forma lógica. En este sentido, este requisito restringe los posibles vehículos de

los pensamientos.

Por otra parte, algunos autores, como Murphy (2002) y Machery (2009), consideran que las inferencias inductivas forman parte de los requisitos explicativos. Según estos autores, la capacidad para anticipar características que tendrá un elemento si pertenece a determinada categoría forma parte de los procesos inductivos. Así, los procesos inductivos implican o bien la generalización de ciertas características de una categoría a elementos nuevos pertenecientes a la misma (un ejemplo de esto sería mi capacidad para anticipar que si planto un árbol va a necesitar sol y agua para sobrevivir), o bien la generalización de propiedades de una categoría a otra categoría (por ejemplo, podemos inferir que los perros tienen pulmones a partir del hecho de que los gatos los tienen). La inclusión de un requisito como éste en una teoría de conceptos apunta a explicar cómo contribuyen los conceptos a este tipo de procesos.

En general, los estudios psicológicos de los procesos de inducción se han concentrado en la influencia de dos variables. Por un lado, han intentado determinar cómo influye la tipicidad de los ítems en la anticipación de propiedades. Por ejemplo, se observó una mayor disposición a afirmar que si los petirrojos (que son pájaros típicos) tienen una determinada enfermedad es probable que todos los pájaros la tengan, que a afirmar que si los patos (que son pájaros atípicos) tienen una enfermedad es probable que todos los pájaros la tengan. Por otro lado, se ha analizado la influencia de la similitud entre dos categorías para la proyección de las propiedades de una a la otra. En este sentido, se ha observado una mayor disposición a atribuir, por ejemplo, las propiedades de los petirrojos a los gansos que las propiedades de las águilas a los gansos (Rips, 1975).⁷

Publicidad

Según este requisito los conceptos deben ser compartibles, no sólo por distintas personas, sino también por un individuo en distintos momentos. Los conceptos deben ser públicos en el sentido de ser accesibles a distintos individuos. Este requisito se postula fundamentalmente por dos razones. En primer lugar, si se pretende que una teoría de conceptos contribuya a una psicología que formule explicaciones intencionales, i.e. que apele a creencias y deseos para explicar el comportamiento de los individuos, estas explicaciones deben permitir generalizaciones que abarquen distintos individuos. Si los conceptos no fueran compartidos, esto no sería posible. Por ejemplo, se puede explicar que Julia agarró una llave

⁷ Para una revisión de los distintos estudios llevados a cabo y los efectos registrados véase Murphy (2002).

y se acercó a la puerta de su casa a partir de su deseo de abrir la puerta y sus creencias de que la puerta estaba cerrada y esa era la llave para abrir esa cerradura.

Este tipo de explicación apela a actitudes proposicionales. Cuando un individuo está en un estado de actitud proposicional, hay un evento que relaciona a ese individuo de un modo particular (que puede ser e.g. de creencia, deseo o temor) con el contenido de la proposición. Por ejemplo si Julia cree que la puerta está cerrada, Julia tiene en su “caja de creencias” una representación mental con el contenido *la puerta está cerrada* (Fodor, 1999). Así, el proyecto de una psicología de este tipo que pretende explicar la conducta por medio de generalizaciones respecto de actitudes proposicionales de los individuos, debe asumir que las actitudes proposicionales de distintos individuos pueden involucrar los mismos contenidos, i.e. estar constituidas por los mismos conceptos. En este caso, es necesario postular que distintos individuos poseen instancias del mismo tipo LLAVE, PUERTA y CERRADURA.

En segundo lugar, muchas teorías lingüísticas apelan a conceptos para explicar la comunicación lingüística, afirmando que asociamos los mismos conceptos a las palabras del lenguaje natural. Si los conceptos no fueran públicos, entonces los pensamientos asociados a las emisiones lingüísticas serían diferentes para cada persona y esto tornaría difícil la comunicación. Puesto que la comunicación de hecho ocurre, esta es una razón para creer que los conceptos son compartidos (Fodor 1999, Prinz 2002).

Como señala Prinz (2002), ciertos casos parecen constituir contraejemplos al fenómeno de la publicidad. Por ejemplo, los niños pequeños y los adultos no parecen compartir los conceptos de entidades biológicas (Keil 1989a) y se puede cuestionar si personas con creencias radicalmente distintas comparten determinados conceptos (Kuhn, 1962). Como afirma Prinz (2002, p. 16) es “precisamente en esos casos en los que los argumentos de la comunicación y la explicación intencional resultan menos persuasivos”. Al considerar a la publicidad como un requisito innegociable, Fodor se concentra en los casos menos “exóticos”, que son los que resultan de interés para las generalizaciones intencionales:

Si el concepto AGUA de todos los demás es diferente del mío, entonces, es literalmente verdadero que sólo yo quise alguna vez un trago de agua, y que la generalización intencional 'las personas sedientas buscan agua' se aplica sólo a mí. [...] Prima facie, parecería que cualquier relativismo conceptual cabal excluiría las generalizaciones intencionales con algún poder explicativo muy serio. (1999, p. 53)

Si bien los casos de diferencias entre niños y adultos o de inconmensurabilidad kuhniana podrían debilitar el requisito de publicidad, no son casos de conceptos privados, en el sentido de no accesibles, sino más bien casos en los que distintos individuos o grupos han adquirido conceptos distintos. Esto no significa que sus conceptos no sean compartibles, de modo que no parece ser evidencia contra los casos ordinarios de comunicación y de conductas que se pueden subsumir en explicaciones intencionales.

Este requisito se relaciona estrechamente con los requisitos de intencionalidad y composicionalidad. El hecho de postular que los conceptos son compartibles, implica que su contenido debe ser compartible por distintos individuos y por un mismo individuo en distintos momentos. Ahora bien, si el contenido de un concepto estuviera determinado por las relaciones inferenciales que mantiene con otros conceptos, tal como afirma la semántica del rol conceptual, la publicidad se vería amenazada. Esto es así porque las relaciones inferenciales entre conceptos podrían llegar a abarcar sistemas de creencias completos y no hay un modo de determinar qué inferencias son constitutivas del contenido y cuáles no (Fodor & Lepore, 1992). Si la determinación del contenido de un concepto depende de su rol en el sistema completo de creencias de un individuo, este sistema es holista. En este sentido, se distingue de un sistema atomista, en el que la determinación del contenido de cada concepto no depende de sus relaciones con ningún otro, y, en menor medida, de un sistema localista, en el cual la determinación del contenido depende de las relaciones con algunos conceptos, pero no con el sistema de creencias completo.⁸

Pues bien, si un sistema es holista, la publicidad se ve amenazada, dado que es poco probable que dos individuos compartan sus sistemas completos de creencias. Incluso es poco probable que un individuo mantenga un sistema de creencias estable sin modificaciones a lo largo del tiempo. Así, en un mismo individuo, el contenido de los conceptos se vería modificado por cada cambio en cualquier parte del sistema de creencias. Esta inestabilidad implica que los conceptos no serían compartibles y, en consecuencia que no serían posibles las generalizaciones intencionales para explicar la conducta. Asimismo, este tipo de propuesta no estaría en condiciones de explicar la comunicación lingüística, que sería un

⁸ Como veremos en los capítulos 2 y 3, las teorías que proponen que los conceptos poseen una estructura semántica interna se comprometen implícitamente con una semántica de rol conceptual, dado que la individuación del contenido dependerá, por lo menos en parte, de la relación del concepto con sus rasgos constitutivos. Un modo de evitar la consecuencia holista que se seguiría de esta tesis, es defender que es posible delimitar un conjunto de inferencias constitutivas del contenido y que, en este sentido, una propuesta de semántica de rol conceptual puede ser localista (Keil & Wilson 2000, Weiskopf 2009b). Me ocuparé de esto en mayor detalle en los próximos capítulos.

misterio. Retomaré esta cuestión en el capítulo 3, en relación a las dificultades de la teoría-teoría para dar cuenta de la publicidad, y en el capítulo 7, donde sugeriré qué estrategias puede adoptar el pluralismo amplio para evitar el holismo.

Por otra parte, el requisito de que los conceptos sean compartibles también está vinculado con el requisito de composicionalidad. Si la posibilidad de formular generalizaciones intencionales depende de la posibilidad de que distintos individuos compartan actitudes proposicionales, entonces la combinación de los conceptos que conforman el contenido de cada actitud proposicional también debe exhibir cierta estabilidad. Para esto, el contenido de las actitudes proposicionales debe poder reducirse a los contenidos de los conceptos constituyentes y las reglas de combinación sintáctica, y, para esto, los conceptos deben ser composicionales.

Así, el análisis de los distintos requisitos revela relaciones entre ellos, por las cuales formular una explicación para algunos de ellos conlleva compromisos implícitos respecto de otros.

3. Requisitos fundamentales de una teoría de conceptos

Machery (2009) distingue entre las teorías de conceptos que se ocupan de los fenómenos vistos en §1 y aquellas que se ocupan de los vistos en §2 e identifica las primeras como teorías filosóficas y las segundas como teorías psicológicas. Machery sostiene que las teorías filosóficas y psicológicas de conceptos tienen objetivos completamente diferentes y, por lo tanto, usan el término “concepto” de maneras muy distintas. Si bien, como sostiene Machery, algunas teorías filosóficas y psicológicas de conceptos se concentran en distintos fenómenos, la distinción que establece resulta demasiado fuerte, puesto que sugiere que no hay relación entre las teorías filosóficas y psicológicas, o incluso que los filósofos y los psicólogos que trabajan sobre conceptos estudian fenómenos radicalmente distintos. Considero que la diferencia entre sus objetivos no conlleva un radical cambio de tema, sino que es más bien una diferencia de énfasis o, en términos de Edwards (2010), una diferencia en el punto de partida.

La distinción que propone Machery parece ignorar las relaciones que hay entre los distintos requisitos que, como sostuve en §2, en muchos casos hacen que sea imposible dar cuenta de uno de ellos sin asumir compromisos (aunque sea de manera implícita) respecto de otros. Así, las teorías que priorizan los fenómenos relacionados con los procesos cognitivos, que Machery identifica como teorías psicológicas, si bien no se ocupan de formular teorías del contenido, ni se comprometen explícitamente con alguna de las teorías disponibles,

presuponen determinadas respuestas a problemas relacionados con el contenido, de los cuales se han ocupado tradicionalmente los filósofos. De la misma manera, si bien algunas teorías que se concentran en el poder expresivo del pensamiento, que Machery identifica como teorías filosóficas, no tienen en cuenta fenómenos relacionados con procesos cognitivos como la categorización, presuponen que los conceptos son el *output* de cualesquiera sean los mecanismos de categorización y, en consecuencia, deben ofrecer un modo de articular los procesos cognitivos con el tipo de representaciones que postulen como conceptos.

Por otra parte, la distinción que propone Machery parece ignorar también el carácter interdisciplinario característico de la investigación en ciencia cognitiva, por el cual el abordaje de los fenómenos cognitivos suele involucrar tanto el punto de vista filosófico como el psicológico. Tomados en conjunto, los fenómenos detallados en §1 y §2, están relacionados por un lado con el poder expresivo de nuestro pensamiento, i.e. con nuestra capacidad representacional, y, por otro lado, con ciertos procesos cognitivos superiores.⁹ El enfoque representacionista, sostiene precisamente que las representaciones mentales cumplen un rol explicativo doble. Por un lado las representaciones mentales contribuyen a la explicación de la relación entre nuestra mente y el mundo, por ser las portadoras de contenido. Por otro lado, las representaciones mentales deben tener características tales que los procesos cognitivos superiores puedan operar sobre ellas. Así, la integración de estos requisitos resulta natural para este enfoque. Las características de los conceptos estarán restringidas o delimitadas tanto por nuestra capacidad representacional, i.e. por los fenómenos de contenido, alcance y composicionalidad, como por los procesos en los que los conceptos intervienen, fundamentalmente, por los fenómenos de categorización y razonamiento.

Actualmente, diversos autores adoptan la mayoría de los requisitos detallados en §1 y §2. Así como algunos filósofos proponen tomar en cuenta los requisitos tradicionalmente psicológicos (Prinz 2002, Weiskopf 2009a), también algunos psicólogos proponen un acercamiento a los requisitos típicamente filosóficos. Como ejemplo de estos últimos, se pueden señalar las consideraciones de Carey y de Markman:

Por supuesto que cualquier teoría del origen de los conceptos requiere alguna idea acerca de qué son los conceptos y cómo se determina su contenido, así

⁹ Adopto aquí la distinción entre procesos cognitivos superiores e inferiores, tal como la reconstruye Machery (2009). Los procesos inferiores incluyen las competencias perceptivas y motrices, mientras que los procesos superiores no toman estímulos perceptivos como *input* (tales como la activación de los conos en la retina) ni emiten *outputs* motores.

como cualquier teoría del contenido conceptual debe acordar con nuestra mejor explicación de cómo se adquieren los conceptos. (Carey 2009, p. 5)

La mayor parte de la investigación en adultos se concentra en la estructura del conocimiento adulto, pero no en su contenido. Si bien esta estrategia ha permitido avances impresionantes en nuestra comprensión del procesamiento cognitivo, en muchas áreas de investigación hemos alcanzado los límites de lo que aprenderemos acerca del procesamiento cognitivo sin al menos cierta atención al contenido del conocimiento adulto. (Markman 2011, p. 141)

Como ilustran estos ejemplos, la división entre teorías filosóficas y psicológicas de conceptos no parece recoger los objetivos explicativos de muchos de los investigadores que trabajan con conceptos. En la misma línea que Carey y Markman, y a diferencia de lo que sostiene Machery, considero que una teoría de conceptos debe responder tanto a cuestiones relacionadas con la metafísica de los conceptos como a cuestiones psicológicas relacionadas con lo que se hace con ellos. Si se tiene el objetivo de incorporar una teoría de conceptos a una teoría de la arquitectura cognitiva, deben tenerse en cuenta ambos aspectos de los conceptos. Consideraré como requisitos fundamentales para una teoría de conceptos la intencionalidad, el alcance y la composicionalidad, que están ligados al poder expresivo de nuestro pensamiento, asimismo, los requisitos de adquisición y categorización que se relacionan con los procesos mentales y, por último, la publicidad, por su rol fundamental en la posibilidad de explicar la comunicación lingüística.

En particular, el requisito de *intencionalidad* resulta fundamental si se pretende articular una teoría de conceptos con una explicación del contenido de los estados mentales. Si se considera los estados mentales como estados intencionales que tienen un contenido y los conceptos como los elementos que, en tanto constituyentes de los pensamientos, son los portadores primitivos del contenido intencional, entonces es necesario incluir una explicación de la semántica de los conceptos. De modo que no sólo es necesario explicar qué clase de entidades son los constituyentes de los pensamientos, sino también en qué consiste su contenido y cómo lo adquieren. Al evaluar la explicación del contenido intencional que ofrece cada teoría, consideraré también al contenido cognitivo, puesto que, como vimos en §1, el contenido intencional no resulta suficiente para la individuación de conceptos.

También incluyo los requisitos de *alcance* y *composicionalidad* puesto que son los que demandan una explicación de nuestra capacidad representacional. Una buena teoría de

conceptos contribuirá a la explicación del pensamiento si puede dar cuenta de la amplia variedad de conceptos que podemos tener. La composicionalidad, a su vez, se propone como la mejor explicación para la productividad y sistematicidad del pensamiento. Estos fenómenos empíricos reflejan la naturaleza del sistema representacional con el que contamos que, no sólo exhibe relaciones estructurales entre los pensamientos, sino que siendo finito nos permite generar una cantidad infinita de pensamientos. La teoría de conceptos que se proponga para explicar el pensamiento debe poder dar cuenta de estos fenómenos, de modo que debe satisfacer el requisito de composicionalidad (aún si se entiende a éste en sentido débil). Por otra parte, dado que la explicación del razonamiento es uno de los objetivos tanto de la filosofía de la mente como de la psicología cognitiva, el requisito de composicionalidad resulta inevitable puesto que, como vimos en §1, contribuye a la explicación del razonamiento deductivo. Con respecto a la *publicidad*, este requisito debe incluirse si se quiere articular una teoría de conceptos con una explicación de la comunicación lingüística, y si se pretende contribuir a una explicación del comportamiento de los individuos que apele a generalizaciones sobre los contenidos de los estados mentales. Incluyo el requisito de publicidad, puesto que considero a ambos proyectos promisorios.

Respecto de los requisitos relacionados con los procesos cognitivos, el requisito de *adquisición* es necesario, puesto que una teoría que postule características ontológicas de los conceptos y cómo estos se utilizan en diversos procesos, resultará insuficiente si no está acompañada de alguna hipótesis plausible acerca de cómo nuestra mente llega a equiparse con esas entidades. Además, aunque no me ocuparé en detalle de los fenómenos asociados al requisito de cambio conceptual, muchos de ellos podrían explicarse por los mecanismos que se postulan en relación a la adquisición. Por otro lado, el requisito de *categorización* es fundamental dado que es un proceso que explica gran parte de nuestro comportamiento hacia los objetos con los que interactuamos. Asimismo, la amplia gama de efectos observados en torno a este proceso (los efectos de tipicidad, de ejemplares y causales, que vimos en §2) constituye una fuente de evidencia, aunque indirecta, respecto de cómo está estructurada la información que almacenamos acerca de las categorías. Ya sea que se considere que esta información es parte constitutiva de los conceptos o no, resulta relevante tener en cuenta estos fenómenos si se pretende articular una teoría de conceptos con una teoría general de la arquitectura cognitiva, para lo cual será preciso determinar el vínculo entre los conceptos y la información estructurada acerca de las categorías.

Por último, algunos fenómenos relacionados con la inducción pueden abordarse a partir de un tratamiento de la categorización, tal como sugiere Murphy (2002). La predicción

de los atributos que tendrá una instancia nueva por pertenecer a una categoría se puede explicar por medio de los mismos procesos que explican la categorización, en tanto involucra “la extensión de información de una categoría a un objeto o categoría nuevos” (Murphy 2002, p. 243). Como mencioné en §2 en relación al requisito de razonamiento, se han llevado a cabo numerosos estudios que intenta relacionar la disposición a atribuir una propiedad de una categoría a una instancia o a otra categoría, con variables de la estructura de las representaciones de las categorías en cuestión. Así, se observó una correlación entre la similitud entre las dos categorías, y la disposición a transferir propiedades entre ellas. Asimismo, se registró un efecto de tipicidad, por el cual la disposición a transferir propiedades de una categoría a otra aumenta cuanto más típica resulta la primera respecto de alguna categoría de segundo orden que las abarque a ambas, como en el ejemplo de las distintas especies de pájaros mencionadas en §2 (Rips 1975). El objetivo de este tipo de estudios es detectar, en tareas de inducción, efectos similares a los observados en tareas de categorización, para así inferir que en ambas intervienen procesos y estructuras del mismo tipo. En este sentido, el estudio de algunas variantes de los razonamientos inductivos, puede abordarse a partir del estudio de la categorización.

De este modo, la adopción de esta lista de requisitos, permite abarcar la mayor parte de los fenómenos que han ocupado tanto a filósofos como a psicólogos y refleja la tendencia creciente a integrar los avances que se han producido desde distintas áreas de la ciencia cognitiva (Prinz 2002, Weiskopf 2009a, Carey 2009). Esta selección constituye un conjunto de requisitos fundamentales, pero esto no significa que el resto de los requisitos mencionados en §1 y §2, en particular las inferencias deductivas e inductivas y el cambio conceptual, de los que no me ocuparé en profundidad, no sean relevantes. Al igual que Weiskopf (2009a), respecto de su propia lista de *desiderata*, considero que son las dificultades para satisfacer este conjunto de requisitos fundamentales las que llevan a postular un pluralismo de conceptos. En este sentido, el agregado de otros requisitos no implicaría diferencias sustanciales ni afectaría los argumentos a favor del pluralismo.

Por otra parte, es necesario aclarar en qué medida las teorías tienen que dar cuenta de estos requisitos. El hecho de seleccionar este conjunto de requisitos explicativos fundamentales implica que una teoría adecuada de conceptos tiene que atribuir a los conceptos características que permitan explicar estos fenómenos o, por lo menos, que sean compatibles con las mejores explicaciones disponibles para ellos. Tal como sostiene Weinberg (2003), una buena teoría de conceptos debería satisfacer ciertos requisitos explicativos, pero esto no implica que todas las explicaciones de esos requisitos deban

basarse en las características metafísicas de los conceptos. Sería una interpretación demasiado fuerte de los requisitos exigirles a las teorías de conceptos que den cuenta de todos estos fenómenos apelando únicamente a las características constitutivas de los conceptos. Sin embargo, sí parece adecuado que las caracterizaciones que ofrecen de los conceptos deban ser compatibles con las mejores explicaciones de estos fenómenos. De este modo, adoptar el conjunto de requisitos fundamentales mencionado no implica descartar apresuradamente teorías que no se ocupan explícitamente de tratar alguno de ellos, sino que propone evaluar si se pueden articular las características que cada teoría propone para los conceptos con explicaciones de cada uno de estos fenómenos. Así, por ejemplo, incluir el requisito de contenido no descarta los enfoques de prototipos, ejemplares o teoría-teoría que no ofrecen una teoría semántica, puesto que generalmente, las teorías de estos enfoques presuponen que los conceptos tienen un contenido representacional (Edwards 2010). Así, aunque no se ocupan explícitamente de formular o adoptar una teoría semántica, no sólo no son incompatibles con la afirmación de que los conceptos tienen contenido, sino que lo presuponen.

De la misma manera, incluir el requisito de categorización no implica descartar al atomismo informacional, que no ofrece una explicación de los procesos de categorización. A partir de una distinción entre las características metafísicas de los conceptos y los procesos cognitivos en los que estos intervienen, Fodor (1999) sostiene que la explicación de cómo se usan los conceptos excede los límites de una teoría de conceptos. Según él, las características metafísicas de los conceptos tienen que ser meramente compatibles con aquello que las teorías psicológicas postulan para dar cuenta de nuestra vida mental. Pues bien, para evaluar si se da o no esta compatibilidad que Fodor mismo concede, es preciso tener en cuenta, entre otros, los fenómenos observados en los procesos de categorización.

Por otra parte, si las mejores teorías psicológicas apelan a representaciones internamente estructuradas para explicar la categorización, entonces para establecer si los conceptos atómicos que Fodor postula son compatibles con las explicaciones psicológicas, hace falta alguna explicación acerca de cómo se relacionan los conceptos atómicos y las entidades estructuradas. Keil, quien ha realizado numerosos estudios en relación a la representación de clases biológicas y sostuvo tradicionalmente que éstas incluyen conocimiento teórico, ha señalado recientemente:

No estoy seguro de la medida en que las facetas potencialmente híbridas de los conceptos son de hecho partes de los conceptos *per se*, o si deben en cambio

considerarse asociadas a los conceptos que, mucho más simples, son en sí mismos átomos en el modo descrito por Fodor [1999]. La ciencia cognitiva aún no ha respondido completamente las razones de Fodor para dudar que muchos fenómenos cognitivos asociados con los conceptos reflejan rasgos estructurales internos de los conceptos en oposición a aspectos de cómo usamos y trabajamos con los conceptos. (Keil 2010, p. 216)

Tomando en cuenta la observación de Keil y las consideraciones en torno a los requisitos, no tiene que asumirse que todo aquello que contribuya a explicar los requisitos fundamentales resulta de este modo constitutivo de los conceptos. Parte de la tarea de una teoría de conceptos es determinar qué contenidos resultan constitutivos de los conceptos y cuáles no, así como qué tipo de relación mantienen los conceptos con la información que, aunque no forme parte constitutiva, está asociada a ellos.

Por último, no todas las teorías que reconstruiré y evaluaré en los próximos capítulos ofrecen un tratamiento explícito de cada uno de los requisitos fundamentales. Así, como he mencionado, las teorías de ejemplares no se ocupan de la publicidad y el atomismo informacional no se ocupa de la categorización. Sin embargo, sugeriré a partir de las tesis principales de cada una de estas teorías cómo podrían enfrentarse a cada uno de los requisitos fundamentales para evaluar si podrían dar cuenta de todos ellos.

Capítulo 2

Tipos de estructura interna.

Primera parte: definiciones, prototipos y ejemplares

En este capítulo y el próximo reconstruiré los debates entre las principales propuestas de estructura conceptual interna. En §1, me ocuparé de la teoría clásica de las definiciones y las objeciones que enfrentó. Según esta teoría, los conceptos consisten en un conjunto de condiciones necesarias y suficientes para pertenecer a una categoría. Esta concepción recibió dos objeciones fatales, que llevaron a su abandono: por un lado, se cuestionó la realidad psicológica de las definiciones y, por otro lado, numerosos estudios reportaron una serie de efectos incompatibles con el uso de definiciones en tareas de categorización. En §2 y §3 me ocuparé de dos enfoques que surgieron como alternativas a las definiciones, el enfoque de prototipos (§2) y el enfoque de ejemplares (§3). Ambos proponen que los conceptos exhiben estructuras en grados, i.e. que hay instancias más y menos típicas de las categorías, en función de su similitud con el resto de las instancias de la categoría. La diferencia principal radica en que el enfoque de prototipos propone que los conceptos están estructurados a partir de información general de la categoría que representan y que para cada categoría disponemos de un conjunto de rasgos típicos que constituyen un prototipo, mientras que el enfoque de ejemplares propone que cada categoría está representada por un conjunto de instancias particulares, constituida cada una por un conjunto de rasgos.

Al reconstruir cada enfoque, me ocuparé puntualmente de sus principales ventajas y dificultades, con el objetivo de establecer, por un lado, qué argumentos conceptuales y qué evidencia motivan su incorporación en un enfoque pluralista de conceptos y, por otro lado, qué limitaciones enfrentan, las cuales un enfoque pluralista deberá resolver. En §4 presentaré una crítica que se ha sostenido respecto de cualquier teoría que postule que los conceptos presentan una estructura semántica interna, la cual afectaría a todos los enfoques presentados en este capítulo. Por último, en §5 señalaré en qué sentido combinar definiciones, prototipos, ejemplares y teorías en una propuesta pluralista puede resultar ventajoso para dar cuenta de los requisitos fundamentales.

1. La teoría clásica de las definiciones

Muchas de las teorías contemporáneas, entre ellas el enfoque de prototipos y el de ejemplares, surgieron por oposición a esta teoría. Según la teoría de las definiciones, los conceptos están estructurados a partir de rasgos definicionales, i.e. condiciones necesarias y

suficientes para caer bajo un concepto.¹⁰ Un ejemplo muy utilizado es el del concepto SOLTERO, que según este enfoque estaría compuesto por los rasgos necesarios y en su conjunto suficientes HOMBRE, ADULTO y NO CASADO. Como aclara Machery (2009), este enfoque no supone que los individuos conocen las condiciones reales de pertenencia a una categoría, sino que afirma que los individuos conciben las condiciones de pertenencia a cada categoría como un conjunto de condiciones necesarias y suficientes.

Los orígenes de este enfoque se remontan a la antigüedad, a la tesis platónica de que los conceptos son esencias abstractas que definen las características de cada categoría, de allí que se lo denomine enfoque *clásico*. Más recientemente, este enfoque puede rastrearse, por un lado en estudios psicológicos en torno a tareas de adquisición de conceptos y categorización llevados a cabo entre 1900 y 1960 (Machery 2007, 2009). Por ejemplo, los estudios de Fischer (1916), Hull (1920) y Smoke (1932) trabajan con categorías bien definidas, en las cuales todos los miembros satisfacen una serie de condiciones necesarias y suficientes. En este sentido, parecen presuponer que los conceptos tienen una estructura definicional.¹¹ Por otro lado, pueden mencionarse ciertos trabajos de semántica léxica en lingüística y psicolingüística, enfocados en determinar las estructuras semánticas asociadas a las palabras del lenguaje natural. En este sentido pueden mencionarse las propuestas de Katz y Fodor (1963) y Katz y Postal (1964).

1.1 Las ventajas de la teoría de las definiciones

Esta teoría resulta especialmente adecuada para explicar algunos de los requisitos fundamentales. En particular, permite dar cuenta del contenido y la composicionalidad, y ofrece una explicación posible para la categorización (aunque, como veremos en la próxima sección, no consigue dar cuenta de ciertos efectos relacionados con esta última). Con respecto al contenido, la teoría de las definiciones ofrece un criterio de individuación adecuado: un concepto refiere a aquellas instancias que satisfacen el conjunto de condiciones necesarias y suficientes que lo constituyen. De la misma manera, la estructura definicional resulta adecuada para dar cuenta de la composicionalidad, puesto que permite postular que los

¹⁰ Para evitar un círculo de definiciones, esta teoría debe asumir que por lo menos algunos conceptos no tienen una estructura definicional, sino que son primitivos.

¹¹ Cabe señalar, sin embargo, que estos estudios no trabajan con clases naturales, sino con categorías artificiales especialmente diseñadas para los estudios. Los miembros satisfacen ciertas condiciones para pertenecer a las categorías, y hay una definición de cada categoría que los sujetos pueden deducir a partir de los miembros presentados y de sus respuestas correctas e incorrectas en las pruebas. En este sentido, cabe la posibilidad de que estos autores no asumieran que todos los conceptos tienen una estructura definicional. Sin embargo, puesto que pretendían revelar ciertas características de la estructura de las representaciones involucradas en los procesos de aprendizaje de categorías y clasificación parece adecuado agruparlos bajo este enfoque.

conceptos compuestos heredan las definiciones de cada uno de sus conceptos constituyentes. Así, por ejemplo, el contenido de TRIÁNGULO EQUILÁTERO, estará exhaustivamente determinado por la unión del conjunto de condiciones que constituyen a TRIÁNGULO (POLÍGONO DE TRES LADOS) y el conjunto de condiciones que constituyen a EQUILÁTERO (DE LADOS CONGRUENTES). De este modo, el contenido de TRIÁNGULO EQUILÁTERO, será POLÍGONO DE TRES LADOS CONGRUENTES.

Con respecto a la categorización, la teoría de las definiciones ofrece una explicación del proceso de categorización. Según este enfoque, el proceso de categorización de una instancia consiste en comparar las características de esa instancia con las condiciones constituyentes de un concepto. Si la instancia satisface todas las condiciones, pertenece a la categoría, si falla aunque sea en alguna de ellas, entonces no. Retomando el ejemplo de TRIÁNGULO, si un ítem no tiene tres lados, entonces no es un triángulo. Si tiene tres lados, pero no es un polígono, i.e. si no es una figura cerrada, tampoco. Para caer bajo la categoría de los triángulos, debe cumplir con las dos condiciones necesarias y en su conjunto suficientes. Esta explicación, parece un modo natural de dar cuenta de nuestra capacidad para categorizar instancias (Margolis & Laurence 2003). Sin embargo, como veremos en la próxima sección, numerosos efectos reportados en la literatura sugieren que no disponemos de definiciones para la mayoría de los conceptos y que, en consecuencia, los procesos de categorización deben incluir operaciones distintas a la verificación de un conjunto de condiciones necesarias y suficientes.

1.2 Las objeciones a la teoría de las definiciones

La principal objeción a la teoría de las definiciones señala la falta de realidad psicológica de ese tipo de estructuras. Si se piensa en ejemplos de conceptos geométricos como TRIÁNGULO, o conceptos de parentesco como ABUELA, que podría definirse como la madre de una madre o un padre, puede parecer intuitivo pensar que los conceptos tienen definiciones. Pero un análisis más profundo revela que esto no es posible para la mayoría de los conceptos. Si se quisiera definir, por ejemplo, un concepto bastante concreto como MESA, no resulta en absoluto evidente cómo definirlo de manera que, por un lado, sea distinto de ESCRITORIO o MESADA y que, por otra parte, no excluya mesas de una sola pata ni tampoco mesas que no tienen patas, como podrían ser mesas en forma de cubo. Así, no podemos especificar un conjunto de condiciones necesarias y suficientes para que algo caiga bajo el concepto MESA. Y si no es posible especificar ese conjunto para la mayoría de los conceptos, i.e. si la mayoría de los conceptos no tienen definiciones, no es posible

sostener que los conceptos son definiciones. En este sentido, distintos autores han señalado las dificultades para identificar un conjunto de condiciones necesarias y suficientes para caer bajo un concepto. El famoso análisis de Wittgenstein (2002) del concepto JUEGO ejemplifica las dificultades para encontrar condiciones necesarias.¹² Para cada característica que supongamos asociada a los juegos, se puede encontrar un contraejemplo que no la tiene. No todos los juegos requieren más de un jugador, no todos tienen reglas preestablecidas, no todos implican competencia, ni tienen ganadores y perdedores, no todos implican actividad física, ni ingenio. Wittgenstein caracteriza aquello que agrupa a los distintos juegos como un “parecido de familia”. Esta noción, como veremos en 2, será retomada por el enfoque de prototipos para caracterizar la relación entre las distintas instancias de una categoría.

Una dificultad análoga surge al intentar formular condiciones suficientes para la pertenencia a un concepto. Fodor (1981) analiza como ejemplo el concepto PINTAR. Considera en primer lugar una definición de Miller, quien define “*x* pinta *y*” como “*x* cubre *y* con pintura”. Fodor señala que estas condiciones no son suficientes, puesto que incluyen casos que no aceptaríamos como casos de pintar, por ejemplo el caso de una fábrica de pintura que explota y cubre con pintura a la gente que se encuentra cerca. Este no es un caso en el que la fábrica ni la explosión pinten a la gente. Podría restringirse la definición incluyendo la condición de que quien pinta sea un agente. Sin embargo, si alguien patea un balde de pintura y cubre el piso con la pintura, este tampoco es un caso de “*x* pinta el piso”. Por más que se agregue a la definición la condición de que la acción de pintar sea la intención primaria de la acción, siguen encontrándose casos que sugieren que esas condiciones no son suficientes. Así, encontrar un conjunto de condiciones suficientes resulta ser tan difícil como especificar condiciones necesarias. En este sentido, el enfoque de las definiciones tiene serias dificultades con el requisito de alcance, dado que sólo podría dar cuenta de unos pocos conceptos, como los conceptos matemáticos o los de parentesco.

Por otra parte, más allá de la búsqueda intuitiva (e infructuosa, por cierto) de definiciones para los conceptos, determinados experimentos psicológicos diseñados para estudiar los procesos de categorización, revelan una serie de fenómenos que resultan incompatibles con la tesis de que los conceptos están estructurados como conjuntos de condiciones necesarias y suficientes para la pertenencia a la categoría representada. Es importante señalar que para que los resultados de estos estudios resulten significativos en relación a la estructura conceptual, se debe aceptar como supuesto inicial que los procesos de

¹² Siguiendo la literatura sobre el tema (Laurence & Margolis, 1999) utilizo la notación para los conceptos aunque el análisis de Wittgenstein es del término “juego”.

categorización operan sobre la estructura constitutiva de los conceptos. Estos estudios se utilizan como evidencia indirecta a favor de determinados tipos de estructura interna, bajo el supuesto de que ciertos efectos resultan más o menos compatibles con los distintos tipos de estructuras. Es decir, estos estudios asumen que las observaciones de efectos en el desempeño sirven como evidencia, aunque indirecta, a favor de uno u otro tipo de estructura interna, puesto que categorizar involucra procesar la estructura interna de los conceptos y distintos tipos de estructura interna predicen distintos tipos de efectos observables al llevar a cabo determinadas tareas.

Con respecto a la estructura definicional en particular, numerosos estudios de tareas de categorización revelan efectos difíciles de explicar por la teoría clásica y sugieren que la mayoría de los conceptos léxicos no muestra ese tipo de estructura.¹³ Por un lado, se han observado casos de indeterminación en los juicios de categorización para ciertos conceptos. Por ejemplo, podemos dudar si una alfombra cae bajo el concepto MUEBLE, lo cual no debería pasar si los conceptos tuvieran estructura definicional. No sólo hay evidencia de casos de duda respecto de la pertenencia a una categoría, sino también de inconsistencia en las respuestas de un mismo individuo para el mismo conjunto de estímulos en ocasiones separadas en el tiempo (Hampton 1998, McCloskey & Glucksberg 1978). Usualmente, se interpreta este fenómeno como un indicio de que nuestras representaciones de categorías exhiben vaguedad o tienen límites difusos, i.e. que las categorías no están representadas por un conjunto de condiciones que permitan determinar para cada elemento si pertenece o no a ella.¹⁴

Por otro lado, se han reportado una serie de efectos llamados de tipicidad, en particular la capacidad de establecer un *ranking* de ejemplares más o menos típicos de una categoría, y la correlación entre la velocidad y precisión con las que se categoriza una instancia, la cantidad de rasgos que esa instancia comparte con el resto de los miembros de la categoría y cuán típica se la considera (Rosch & Mervis, 1975). Distintos estudios muestran que los ejemplares más típicos se categorizan más rápido y más precisamente (i.e. con menor tasa de error), que los menos típicos. Y estos efectos exhiben una correlación con la cantidad de atributos que cada ejemplar comparte con el resto de los miembros de la categoría y con el lugar que ocupan en el ranking. Ninguno de estos efectos debería observarse si los conceptos

¹³ A grandes rasgos, los conceptos léxicos se corresponden con palabras (como opuestas a frases) en el lenguaje natural.

¹⁴ Machery (2009) sostiene que el hecho de que existan instancias para las cuales no está claro si son miembros o no de una categoría dada es consistente con la teoría de las definiciones, puesto que las definiciones pueden producir vaguedad si los elementos de la definición son vagos. Sin embargo, si los elementos de una definición son vagos, no parece completamente lícito considerarlos un conjunto de condiciones necesarias y suficientes.

tuvieran una estructura definicional. Me ocuparé de estos efectos en mayor detalle en la próxima sección, en relación al enfoque de prototipos.

Es importante señalar que, si bien estas objeciones resultaron en el abandono de esta teoría como una teoría general de conceptos, algunos autores defienden este tipo de estructura para ciertos conceptos. Este sería el caso, por ejemplo, de los conceptos geométricos, como TRIÁNGULO, o RECTA y los de parentesco (Pinker & Prince 1999). Otra vertiente contemporánea de defensa de las definiciones sostiene que la teoría clásica ha sido malinterpretada y que su alcance pretendido es dar cuenta de las estructuras conceptuales subyacentes a los juicios modales, y relativas a aquello que nos resulta inteligible. Rey (2005) señala que las definiciones son las más adecuadas para dar cuenta de nuestra capacidad para llevar a cabo juicios relativos a qué instancias caerían bajo un concepto en situaciones contrafácticas, fundamentalmente para explicar los juicios acerca de entidades *inexistentes* (como por ejemplo la atribución de mente a los fantasmas). Asimismo, el análisis conceptual clásico permitiría explicar por qué “círculo cuadrado” no nos resulta inteligible. Este tipo de posturas, sin embargo, también reconocen que los conceptos pueden tener otros tipos de estructuras y no constituyen una defensa de la teoría clásica como una teoría general de conceptos.

2. El enfoque de prototipos

El enfoque de prototipos surgió, en la década de 1970, por oposición a la teoría clásica de las definiciones. Este enfoque se opone a la idea de que la pertenencia a las categorías depende de la satisfacción de un conjunto de condiciones necesarias y suficientes y sugiere, en cambio, que la pertenencia a las categorías es una cuestión de grado. La caracterización que ofrece Hampton de un prototipo (a los que se refiere como “conceptos polimorfos”) refleja claramente esta oposición:

Un Concepto Polimorfo puede definirse como uno en el que una instancia es clasificada como perteneciente a una clase determinada, si y sólo si posee por lo menos una cierta cantidad de un conjunto de rasgos, ninguno de los cuales necesita ser por sí mismo necesario ni suficiente. (1979, p. 450)

Esta caracterización es en su mayor parte negativa, en tanto señala que este enfoque niega que los conceptos estén estructurados como definiciones. Con respecto a la caracterización positiva de los prototipos, en uno de los trabajos fundacionales de este

enfoque, Rosch y sus colegas (1976) proponen que el recorte que hacemos del mundo en categorías se basa en la coocurrencia de rasgos. Hay propiedades que, en el mundo, tienden a ocurrir conjuntamente y nuestras categorías (por lo menos las de objetos concretos) recogen las combinaciones de propiedades que más frecuentemente ocurren en conjunto. Así, la idea central detrás de las teorías de prototipos es que los conceptos están estructurados en función de los rasgos que más frecuentemente ocurren en conjunto, que son los rasgos típicos de las categorías. Una consecuencia de esto es que habrá instancias más y menos típicas de una categoría, en función de cuántos rasgos compartan. A diferencia de lo que sostenía la teoría clásica, las instancias de una categoría no comparten un conjunto de condiciones necesarias y suficientes, sino que tienen entre sí más bien cierto “parecido de familia”, i.e. una superposición parcial de rasgos (Rosch & Mervis 1975).

Las teorías de prototipos coinciden en postular que los conceptos consisten en conjuntos de rasgos estadísticamente frecuentes. Sin embargo, se han propuesto distintos modelos teóricos que representan a los prototipos de distintas maneras. El modo más común de representarlos es como conjuntos de representaciones de rasgos ponderados en función de dos dimensiones: la correlación, que es la frecuencia con la que dos rasgos ocurren conjuntamente, y la distintividad o validez indicadora [*cue validity*], que es la probabilidad condicional de que una instancia pertenezca a una categoría dado un determinado rasgo. Y esto último depende de si los rasgos son compartidos con gran cantidad de conceptos o son distintivamente característicos de uno en particular (Rosch, Mervis, Gray, Johnson & Boyes-Braem 1976). Estas dos dimensiones recogen, respectivamente, cuán fuertemente está asociado cada rasgo a una categoría (en función de la frecuencia con la que ocurre en instancias de la categoría) y cuán informativo o útil resulta cada rasgo a la hora de distinguir una categoría de otra.

Estas dos dimensiones de variación deben diferenciarse. Por ejemplo ROJO es un rasgo altamente saliente de las manzanas. Sin embargo, su utilidad diagnóstica es menor, puesto que es un rasgo compartido por muchas categorías alternativas. Justamente, la noción de “validez indicadora” propuesta por Rosch y Mervis (1975), refiere a la utilidad diagnóstica de los rasgos.¹⁵ Se trata de la probabilidad condicional de que una instancia pertenezca a una categoría dado un determinado rasgo. La relevancia de un determinado rasgo para predecir la pertenencia a una categoría dada aumenta con el aumento de la frecuencia con la que ese rasgo está asociado con esa categoría y disminuye con el aumento de la frecuencia con la que

¹⁵ Adoptaré la denominación “utilidad diagnóstica” en lugar de la original de Rosch y Mervis, “validez indicadora”, dado que en castellano la primera resulta más clara.

ese rasgo está asociado con otras categorías distintas. En otras palabras, los rasgos con mayor utilidad diagnóstica son aquellos que resultan más relevantes a la hora de determinar la pertenencia a una determinada categoría, y tal relevancia depende de la frecuencia con la que ese rasgo figura en instancias de esa categoría, así como de la frecuencia con la que figura en otras categorías.

Un modelo que ha resultado especialmente fructífero para la explicación de la categorización y la combinación de conceptos, a pesar de que, como argumentaré en §2.2, enfrenta ciertas limitaciones con respecto a la composicionalidad, divide los rasgos entre atributos y valores. Los valores son los rasgos posibles que pueden exhibir las instancias de esa categoría para cada uno de los atributos. Por ejemplo, el prototipo de MANZANA puede incluir, entre otros, los atributos COLOR y FORMA, y el atributo COLOR, a su vez, puede incluir los valores ROJA, VERDE y AMARILLA. Según este modelo, todos los valores y atributos están ponderados. En el caso de los atributos, los pesos se distribuyen de acuerdo a la utilidad diagnóstica de cada uno, i.e. en función de cuánto contribuyen en la tarea de reconocer elementos pertenecientes a la categoría. Por ejemplo, el atributo COLOR puede no ser de mucha utilidad a la hora de determinar si algo es un perro, mientras que resulta relevante en el reconocimiento de la arena (si de lo que se trata es de diferenciarla, por ejemplo, de tierra o arcilla). En el caso de los valores, los pesos se asignan en función de cuán sobresaliente resulta ese rasgo, esto es, en función de la frecuencia con la que ese rasgo ocurre en instancias de ese concepto o en función de cuán perceptible resulta. En el caso del prototipo de MANZANA, puesto que la mayoría de las manzanas que encontramos son rojas, ROJA tendrá un peso mayor que VERDE y éste, a su vez, uno mayor que AMARILLA (Smith, Osherson, Rips & Keane 1988).

En lo que sigue utilizaré este modelo para ilustrar las ventajas que ofrecen los prototipos fundamentalmente en relación a la categorización, así como las dificultades que enfrentan en relación a los requisitos de composicionalidad.

2.1 Ventajas y evidencia a favor de los prototipos

Postular una estructura de información prototípica tiene, por un lado, la ventaja de explicar satisfactoriamente la adquisición, a partir de la detección de regularidades estadísticas y, por otro lado, determinados efectos observados en tareas de categorización, a partir de una función de similitud que compara rasgos percibidos con los rasgos representados en cada concepto. Con respecto a la adquisición, los defensores de este enfoque

sostienen que a partir de experiencias con instancias de categorías, se produce un registro de regularidades y correlaciones estadísticas entre rasgos que dan lugar a los prototipos.

Con respecto a la categorización, dado el modelo de representación prototípica presentado en §2, Smith, Osherson, Rips y Keane (1988) adoptan una fórmula de Tversky que arroja una medida de similitud computando tanto los rasgos comunes como los no comunes entre dos representaciones. Para aplicar la fórmula, resulta razonable suponer, siguiendo a Smith y sus colegas, que la instancia que se quiere categorizar se representa como un prototipo, i.e. como una estructura de atributos y valores, pero sin pesos diagnósticos asignados a los atributos. El peso diagnóstico que se asignará a cada uno de estos atributos para computar la similitud, dependerá de los pesos diagnósticos de cada atributo en la representación prototípica de la categoría para la que se llevará a cabo la comparación. Cada rasgo (i.e. cada valor), por otra parte, adopta como peso la suma de los pesos de todos los valores para el atributo al que ese valor corresponde.¹⁶ Los autores adoptan la fórmula de contraste de Tversky (1977) para medir la similitud entre una representación *I* de una instancia y un prototipo *P*:

$$(1) \text{ Sim}(I,P) = \sum_i [af_i(P \cap I) - bf_i(P - I) - cf_i(I - P)]$$

En (1), *i* se refiere a un determinado atributo, $P \cap I$ determina el conjunto de rasgos que *I* tiene en común con *P* para ese atributo, $P - I$ señala el conjunto de rasgos distintivos de *P* para el atributo *i* (i.e. rasgos que son característicos del prototipo y no están presentes en la instancia), mientras que $I - P$ señala el conjunto de rasgos distintivos de *I* para el atributo *i* (rasgos presentes en la instancia *I* y no en *P*). Por otra parte, f_i multiplica el peso para cada conjunto de rasgos por el peso diagnóstico del atributo *i*. En el ejemplo, Smith y sus colegas asumen que el valor de *a*, *b* y *c* es 1, para simplificar la ecuación, pero se podrían asignar distintos valores a estas variables para otorgar distinta relevancia al conjunto de rasgos en común y a los conjuntos de rasgos distintivos. Para que la medida de similitud que arrojan este tipo de fórmulas tenga algún significado, es necesario establecer un umbral que las instancias deben superar para caer bajo un determinado prototipo. Además, para normalizar los resultados, los pesos asignados a cada atributo pueden adoptar valores entre 0 y 1, y la

¹⁶ Esta es una simplificación que el modelo asume. Podría ocurrir que, para un determinado atributo, una instancia tuviera más de un valor en común con el prototipo de la categoría con la que se la compara y, en ese caso, no tendría sentido asignar a cada uno la suma total de los valores.

suma de todos los pesos diagnósticos debe ser igual a 1, de manera que la medida de similitud siempre dé un resultado entre 0 y 1.

Este tipo de modelo permite dar cuenta de una serie de efectos observados en relación a la estructura de las representaciones de categorías y al desempeño en tareas de categorización. Como mencioné en §1.1 a propósito de las objeciones a la teoría de las definiciones, se ha observado que, por un lado, podemos evaluar si una determinada instancia es un ejemplar más o menos típico de una categoría, i.e. podemos establecer grados de tipicidad con respecto a una categoría y, por otro lado, el grado de tipicidad está relacionado con la cantidad de rasgos que comparte con el resto de los miembros de la categoría. Por ejemplo, hay una correlación entre el grado de tipicidad de una manzana y una aceituna con respecto a la categoría FRUTA y la cantidad de rasgos que cada una comparte con el resto de los miembros de la categoría: una manzana es una instancia más típica de FRUTA que una aceituna y una manzana exhibe más rasgos típicos de las frutas que una aceituna (Hampton 1979, Rosch 1973, Rosch & Mervis 1975).

Este tipo de resultados han sido considerablemente discutidos. Por un lado, Barsalou (1985; 1987) muestra que los juicios de tipicidad no son uniformes, sino que varían de un individuo a otro, y en un mismo individuo en distintos momentos y contextos. Barsalou considera que la similitud con la tendencia central de una categoría no es la única variable que puede resultar en distintos grados de tipicidad para distintas instancias, sino que hay otros factores, en particular la similitud con ideales (asociados con objetivos o funciones que cumple la categoría), y la frecuencia de instanciación de cada instancia, que también pueden determinar estructuras en grados para las categorías. Así, señala que dependiendo del contexto, un individuo puede considerar como más típico al ejemplar que mejor cumple con un objetivo asociado a la categoría, mientras que en otro, el ranking de tipicidad puede establecerse en función de la similitud con la tendencia central de la misma categoría. En este sentido, el hecho de que nuestros juicios de categorización revelen una estructura en grados para las categorías, no constituye evidencia decisiva a favor de los prototipos.

Sin embargo, como señala Hampton (2006), los defensores del enfoque de prototipos no consideran esta evidencia de manera aislada, sino que destacan la correlación entre los juicios de tipicidad y las variables que se miden en muchos de los estudios de categorización: los tiempos de respuesta y la tasa de error. Los estudios que brindan evidencia más fuerte a favor de las estructuras prototípicas muestran una correlación entre la cantidad de rasgos típicos de una categoría que exhibe una instancia y el tiempo de respuesta para categorizar esa instancia. Las instancias más similares a las instancias típicas se categorizan más rápido

que aquellas menos similares, al mismo tiempo que se cometen menos errores con las instancias más típicas (Hampton 1979, Rips, Shoben & Smith 1973, Rosch & Mervis 1975). Sin embargo, como se verá en §3, también se han observado otros efectos en tareas de categorización que resultan difíciles de acomodar en la perspectiva de que todos los conceptos están estructurados como prototipos.

2.2 Las dificultades del enfoque de prototipos

El enfoque de prototipos tiene dificultades fundamentalmente para satisfacer el requisito de composicionalidad. Se han esgrimido numerosas críticas en relación a este requisito. Las objeciones principales señalan, primero, que hay muchos casos de conceptos complejos sin prototipo y, segundo, que aun en los casos en los que habría un prototipo del concepto complejo que resulta de combinar otros más simples, los rasgos de éste no se obtienen combinando de modo composicional los rasgos de los prototipos constituyentes. Es importante distinguir estas objeciones de las críticas al enfoque de las definiciones. Al señalar que muchos conceptos complejos no tienen prototipo, no se cuestiona la realidad psicológica de los prototipos. Los numerosos efectos de tipicidad mencionados en §2.1 parecen resultar suficientes para afirmar que hay prototipos asociados a las categorías. La cuestión es si estas estructuras prototípicas resultan constitutivas de la estructura conceptual o no (Fodor 1981).

Con respecto a la primera objeción, en ciertos casos, al combinar conceptos que tienen prototipo, el concepto resultante no lo tiene. Fodor propone ejemplos como CIUDADES AMERICANAS SITUADAS EN LA COSTA ESTE UN POCO AL SUR DE TENNESSE y ABUELAS LA MAYORÍA DE CUYOS NIETOS ESTÁN CASADOS CON DENTISTAS (Fodor 1981) y otros ejemplos usando operaciones booleanas como NO GATO (Fodor 1999), ROSA SI ES CUADRADO y NÚMERO NO PRIMO (Fodor & Lepore 1996). Estos conceptos no tendrían una estructura en grados. Por ejemplo, no hay instancias más o menos típicas de NO GATO. Un zapallito y un ventilador son instancias igualmente típicas de NO GATO, porque ninguno es un gato. El hecho de que ciertos conceptos complejos no exhiban estructura prototípica podría ser interpretado como un problema de alcance del enfoque, en tanto sería una limitación en el espectro de conceptos que se pueden representar de este modo. Sin embargo, puesto que en estos casos pueden identificarse los efectos de tipicidad para tareas que involucran a los conceptos que se quiere combinar (por ejemplo, ABUELA, MAYORÍA, NIETO CASADO y DENTISTA), la dificultad es diferente. No se trata de un problema de alcance, sino de si este tipo de estructuras pueden dar cuenta de la productividad y la sistematicidad del pensamiento. Como vimos en el capítulo 1, el requisito

de composicionalidad se propone como la mejor explicación para estos fenómenos. En términos de Fodor (1981, p. 297):

Hay productividad porque las propiedades semánticas se heredan de los términos a las frases (o, de modo equivalente, para nuestros objetivos, de conceptos léxicos a conceptos frasales). A la inversa, nada podría ser el significado de una palabra, a menos que sea aquello que heredan las frases en las que la palabra figura.¹⁷

Así, si al combinar los conceptos de ABUELA, MAYORÍA, NIETO, CASADO y DENTISTA (para los cuales resulta plausible que haya representaciones prototípicas), el resultado no tiene prototipo, entonces los conceptos combinados no están contribuyendo con estructuras prototípicas al concepto frasal resultante. Y si esto es así, argumenta Fodor, entonces las estructuras prototípicas no pueden ser constitutivas de los conceptos, puesto que, si lo fueran, la productividad quedaría sin explicación.

Con respecto a la segunda dificultad, incluso en los casos de conceptos complejos respecto de los cuales no se discute que tengan un prototipo, se puede objetar que ese prototipo no es el resultado de la combinación de otros prototipos siguiendo reglas de composición. Por un lado, al combinar ciertos prototipos pueden surgir propiedades nuevas que no estaban en los prototipos iniciales y, por otro lado, pueden perderse propiedades típicas de los prototipos iniciales (Connolly, Fodor, Gleitman & Gleitman 2007, Fodor 1999, Fodor & Lepore 1996, Kunda, Miller & Claire 1990, Medin & Shoben 1988, Osherson & Smith 1981). El primer fenómeno se conoce como el de las *propiedades emergentes*. Por ejemplo, al listar las propiedades de las cucharas de madera, generalmente se les atribuyen los rasgos de ser grandes y usarse para servir (en lugar de usarse para comer), aunque estos no sean rasgos típicos de las cucharas ni de las cosas de madera.

Por otro lado, hay casos en los que el prototipo resultante no hereda propiedades que son típicas de los prototipos que se combinan. Este es el caso de, por ejemplo, HOMBRE ORQUESTA que es típicamente un hombre con muchos instrumentos (en general un bombo, platillos, una armónica, un acordeón y una trompeta) y cuyo ejemplar típico no se aleja mucho de Bert en *Mary Poppins*. Este concepto complejo hereda de ORQUESTA algo así como el atributo INSTRUMENTOS MÚLTIPLES, pero no el atributo COMPUESTA POR

¹⁷ En el contexto del argumento de Fodor, se asume que los conceptos representan aquello que las palabras expresan y son, en este sentido, representaciones de sus significados.

VARIOS MÚSICOS. Estos contraejemplos se proponen para mostrar que si se trata a los conceptos compuestos como prototipos, su contenido no puede explicarse por combinación de prototipos de otros conceptos más simples, debido a que el requisito de composicionalidad establece que la semántica de los conceptos complejos debe quedar exhaustivamente determinada por los contenidos de los conceptos constituyentes, junto con las reglas de combinación.

Para hacer frente a estas dificultades, se han elaborado modelos que proponen reglas de combinación específicas para los prototipos (Costello & Keane 2000, Smith *et al.* 1988). Por ejemplo, el *Modelo de Modificación Selectiva* de Smith y sus colegas, establece reglas para la combinación de sustantivos y adjetivos. Este modelo supone que el sustantivo ofrece el prototipo más estable para operar y que el adjetivo introduce modificaciones en este prototipo. En el proceso de combinación, cada atributo del adjetivo selecciona el atributo correspondiente en el sustantivo; para los atributos en los que hay correspondencia, aumenta en el sustantivo la saliencia del valor nombrado por el adjetivo y aumenta el valor diagnóstico del atributo en cuestión.

Si bien este modelo permite dar cuenta de algunos casos sencillos de combinaciones de sustantivos y adjetivos, y se ha extendido a combinaciones con adverbios, no consigue dar cuenta de casos más complejos, ni de los problemas señalados relativos a las propiedades emergentes y a las propiedades perdidas en la combinación. En consecuencia, este tipo de modelos enfrentan numerosas objeciones. Por un lado, los casos que permiten explicar son demasiado sencillos e introducen simplificaciones (como la cantidad de atributos en común entre el sustantivo y el adjetivo), que no parecen cumplirse en los casos reales. Así, no parecen aplicables a casos en los que, por ejemplo, dos conceptos no tienen rasgos en común. Las explicaciones de estos modelos, aunque resultan atractivas, siguen sonando un tanto *ad hoc*. En palabras de Murphy (2002), la combinación de prototipos parece ser más interactiva que lo que estos modelos describen, resultando difícil de reducir a un conjunto de reglas.

Por otra parte, los fenómenos de las propiedades emergentes y las propiedades perdidas en la combinación parecen estar relacionados con el conocimiento de instancias particulares y el conocimiento de trasfondo en general. Retomando el ejemplo de las cucharas de madera, el hecho de que sea de madera debería, según el modelo de Smith y sus colegas, modificar sólo un atributo relativo al material. Sin embargo, al combinar estos dos conceptos, surgen propiedades relativas al tamaño y la función, que no son propiedades de las cosas de madera en general (Medin & Shoben 1988). En consecuencia, no parece posible explicar el fenómeno de las propiedades emergentes apelando únicamente a representaciones

prototípicas. El modelo de Costello y Keane (2000), por ejemplo, que resulta más adecuado para dar cuenta de tales propiedades, parte de las representaciones de prototipos, pero incluye también otros tipos de representaciones. En sus palabras:

La gente hace uso de una variedad de conocimiento cuando interpreta frases nuevas. En este sentido, asumimos que el proceso de combinación puede hacer uso de prototipos de los conceptos constituyentes, instancias específicas de esos conceptos, prototipos e instancias de conceptos relacionados, teorías de dominio general y representaciones de eventos específicos que involucran estos conceptos. En pocas palabras, la teoría de la restricción propone que el proceso de combinación tiene acceso directo al contenido total de la memoria. (p. 305)

Así, la defensa de que los prototipos se pueden combinar sin recurrir a información de trasfondo o a representaciones de otro tipo, tales como los ejemplares, parece funcionar para casos simples, pero no para los casos más complejos de combinación, y no consigue hacer frente al tan discutido fenómeno de las propiedades emergentes. Estas propiedades parecen surgir de otros tipos de representaciones, como los ejemplares, que abordaré a continuación, y las teorías y redes causales, de las que me ocuparé en el próximo capítulo.

3. El enfoque de ejemplares

El enfoque de ejemplares (Medin & Schaffer, 1978; Nosofsky, 1986; Nosofsky & Palmeri, 1997) propone que los conceptos son conjuntos de representaciones de ejemplares, i.e. instancias particulares, que caen bajo una misma categoría. Los ejemplares son representaciones de instancias específicas con las que se ha estado en contacto, de modo que, a diferencia de los prototipos, no se basan en información de tipo estadístico. Según las teorías que postulan este tipo de representaciones, cada categoría está representada por medio de un conjunto de representaciones de instancias particulares de la misma. La mayoría de los modelos disponibles asumen que cada ejemplar consiste en un conjunto de representaciones de rasgos.

Es importante señalar que este enfoque no niega la posibilidad de que los conceptos incorporen, además de la información relativa a cada ejemplar del conjunto, información general relativa a la categoría. Sin embargo, en particular para los procesos de categorización, sostienen que la información general de la categoría no interviene, sino que se basan

exclusivamente en los rasgos de cada ejemplar (Medin & Schaffer, 1978). De este modo, este enfoque otorga una relevancia mayor a la información acerca de instancias particulares para los procesos cognitivos conceptuales que a la información categorial general. Esto está vinculado con una de las motivaciones principales del enfoque, que consiste en explicar los efectos que tienen en distintas tareas (fundamentalmente en el aprendizaje y la categorización), nuestros encuentros previos con instancias particulares de las categorías.

Al igual que en el enfoque de prototipos, según los defensores de este enfoque, los procesos cognitivos conceptuales se basan en el cómputo de similitudes. Sin embargo, en el caso de este enfoque se trata de cálculos de similitud entre ejemplares y otras representaciones, sin apelar a una síntesis o abstracción de los rasgos comunes entre los distintos ejemplares. Distintas teorías correspondientes a este enfoque ofrecen distintos modelos de ejemplares. Detallaré aquí brevemente el *Modelo de Contexto* de Medin y Schaffer (1978), uno de los primeros del enfoque, y probablemente el más influyente. Medin y Schaffer proponen que cada ejemplar es una representación estructurada en dimensiones y valores. Como ejemplo, proponen un conjunto de figuras geométricas, estructuradas en cuatro dimensiones: color, forma, tamaño y número. Probablemente con el objetivo de presentar un ejemplo sencillo, sugieren, para cada una de las dimensiones, dos valores posibles: rojo o azul, triángulo o círculo, grande o pequeño y uno o dos, respectivamente. Dado que este ejemplo presenta únicamente dos alternativas, los valores de cada una pueden representarse con 1 y 0. Así, por ejemplo, considerando las dimensiones en el orden presentado, un triángulo azul pequeño puede representarse en notación binaria como 0101, y dos círculos rojos grandes como 1010.

El nombre del modelo alude al rol atribuido al contexto en los procesos de categorización. El contexto de la tarea puede influir en la atención otorgada a cada dimensión en el cómputo de similitudes. Este aspecto se incorpora para dar cuenta de los efectos que puede tener la puesta a prueba de hipótesis en la categorización. Según Medin y Schaffer, los individuos pueden aplicar distintas estrategias durante el aprendizaje de categorías y formular distintas hipótesis respecto de cómo se agrupan los distintos ejemplares, lo cual modifica la atención dirigida a cada dimensión. Por ejemplo, si frente a una serie de estímulos que presentan las dimensiones del ejemplo, los individuos pusieran a prueba la hipótesis de que todos los estímulos azules corresponden a una categoría y los rojos a otra, es probable que presten menos atención a la forma de los estímulos.

3.1 Ventajas y evidencia a favor de los ejemplares

La tesis central de los enfoques de ejemplares con respecto a la categorización señala que estos procesos se basan en la recuperación de información sobre ejemplares almacenada en la memoria y en procesos de cómputo de similitud entre la instancia que se quiere categorizar y estos ejemplares. Al igual que en el enfoque de prototipos, los modelos de categorización elaborados son probabilísticos: la probabilidad de que una instancia i sea identificada como un miembro de la categoría c aumenta directamente en función de la similitud entre la instancia i y los ejemplares almacenados en c , e inversamente en función de la similitud entre i y los ejemplares almacenados asociados con categorías alternativas. El modelo de Medin y Schaffer (1978), en lugar de incorporar una regla de adición para computar la similitud como hacen las teorías de prototipos (que suman los rasgos compartidos), incorpora una regla multiplicativa, que implica que se favorezcan los casos de alta similitud con un ejemplar específico de una categoría, por sobre la similitud moderada con varios ejemplares.

De este modo, consigue dar cuenta de ciertos efectos observados. Contra lo que predicen las teorías de prototipos, muchos estudios indican que podemos categorizar instancias atípicas de una categoría igualmente o más rápido que instancias típicas, en caso de haber tenido experiencias previas con instancias particulares similares al ejemplar atípico en cuestión (Medin & Schaffer, 1978). Estos efectos, relativos a la familiaridad con instancias, se han denominado “efectos de ejemplares”, dada su relación con el conocimiento de instancias particulares de las categorías (Murphy, 2002). Así, la principal ventaja de este modelo es su capacidad para predecir, a la vez, los efectos de tipicidad observados por los defensores de los prototipos (dado que estos se explican por una similitud alta con los ejemplares típicos de la categoría), y los casos en los que se categorizan muy rápidamente instancias atípicas de las categorías, los cuales resultan difíciles de acomodar por los modelos de prototipos (Nosofsky & Zaki, 2002).

Por otra parte, como señala Prinz (2002), las teorías de ejemplares presentan, con respecto a las teorías de prototipos, una ventaja para la representación de categorías de nivel supraordinado, como mueble, herramienta o vehículo.¹⁸ Esto es así porque, mientras resulta

¹⁸ Dado que las instancias pertenecen a múltiples categorías a la vez (e.g. Farofa pertenece a la categoría de los perros, la de los mamíferos y la de los animales), en la literatura de categorización suele aceptarse una taxonomía de categorías. Las categorías se organizan jerárquicamente por su nivel de inclusión y abstracción, formando una taxonomía en la que las categorías de un nivel incluyen a todas las de los niveles inferiores (Murphy, 2002). Distintos estudios revelaron un nivel básico de categorización en el que se ubican las instancias (Rosch *et al.* 1976). Este nivel sería el que utilizamos por *default*, naturalmente, ante encuentros con instancias. Por ejemplo, es más probable que identifiquemos a una instancia como PERRO que como MAMÍFERO o

difícil abstraer la tendencia central de categorías como mueble o vehículo (puesto que es poco plausible que un prototipo pueda recoger tendencias centrales entre los miembros dispares de estas categorías), éstas se podrían representar en cambio fácilmente por medio de un conjunto de representaciones de sus instancias. Esta podría parecer una ventaja sobre los prototipos en relación al requisito de alcance. Sin embargo, como desarrollaré en la próxima sección, este no es el caso, puesto que este enfoque tiene dificultades para dar cuenta de conceptos abstractos y teóricos, con cuyas instancias no estamos en contacto. Sería más preciso afirmar, en cambio, que el enfoque de prototipos y el enfoque de ejemplares tienen alcances distintos, i.e. que cada uno resulta más adecuado para distintos tipos de categorías.

Como ventaja adicional, es importante destacar la relación entre los modelos de almacenamiento y procesamiento de ejemplares, y los modelos de memoria (Nosofsky 1988, Nosofsky, Little, Donkin, Little & Fific 2011, Nosofsky, Cox, Cao & Shiffrin 2014, Nosofsky, Cao, Cox & Shiffrin 2014). Nosofsky y sus colegas extendieron el alcance de los modelos de ejemplares, de modo tal que pudieran dar cuenta de algunos de los principales efectos relativos a procedimientos de búsqueda en la memoria reportados en la literatura. Sin embargo, como sostendré en la próxima sección, este enfoque enfrenta dificultades relativas a algunos de los requisitos.

3.2 Dificultades de este enfoque

Este enfoque tiene, fundamentalmente, dificultades de alcance, dado que no se puede explicar la posesión de conceptos teóricos, por ejemplo ELECTRÓN, como un conjunto de ejemplares (Prinz 2002, Machery 2009). Se pueden encontrar muchos ejemplos de categorías, de las cuales nunca entramos en contacto con sus instancias particulares, y para las cuales, en consecuencia, no disponemos de un conjunto de ejemplares. Esta dificultad podría tratar de resolverse apelando a combinaciones de ejemplares de los que sí disponemos. Sin embargo, las teorías de ejemplares tienen múltiples problemas en relación al requisito de combinación de conceptos, como desarrollaré a continuación.

Con respecto a los conceptos complejos para los cuales disponemos de representaciones de ejemplares en la memoria, este enfoque no tiene dificultades. Así, podría hacer frente al problema de las propiedades emergentes que aqueja al enfoque de prototipos.

CHIHUAHUA. Según Rosch *et al.* (1976), este es el nivel más inclusivo en el que la mayoría de los miembros de la categoría comparten la mayor cantidad de rasgos. Las categorías de niveles menos inclusivos que el básico se denominan *subordinadas* y las de niveles más inclusivos, *supraordinadas*. Así, dos instancias de chihuahua tienen más rasgos en común entre sí que dos instancias de perro, pero la categoría perro incluye más individuos que la categoría chihuahua. Asimismo, dos instancias de mamífero tienen menos rasgos en común que dos instancias de perro, pero la clase de los mamíferos es más inclusiva que la de los perros.

Las teorías de ejemplares podrían explicar algunos de los ejemplos típicos propuestos contra las teorías de prototipos, como CUCHARA DE MADERA, puesto que la emergencia de los rasgos GRANDE y PARA SERVIR, podría proceder, precisamente, de representaciones almacenadas de instancias particulares de cucharas de madera (Prinz, 2002). Sin embargo, las teorías de ejemplares no ofrecen una explicación adecuada de la generación de combinaciones nuevas de conceptos, puesto que no explican cómo podrían combinarse las representaciones de distintas instancias particulares para dar lugar a un nuevo ejemplar, ni cómo se da la combinación de distintos conjuntos de ejemplares. En este sentido, por sí mismas, no son capaces de explicar la productividad ni la sistematicidad del pensamiento, sino que sólo pueden dar cuenta de la activación de representaciones de instancias conocidas.

Por último, este enfoque también tiene dificultades relacionadas con la publicidad, debido a que no resulta plausible que distintos individuos representen los mismos ejemplares, sino que dado que cada uno tiene experiencias con distintos ejemplares, cada uno tendrá conceptos distintos. Esto no es algo que los teóricos de este enfoque nieguen, sino que lo incorporan en los modelos. Así, por ejemplo, Medin y Schaffer afirman que en el contexto de aprendizaje de una categoría pueden ponerse en juego distintas estrategias e hipótesis que conducen a almacenar información sobre distintas dimensiones y a asociar distintas representaciones a cada categoría. Como consecuencia, no compartiríamos conjuntos de ejemplares. Así, este enfoque no puede dar cuenta de la comunicación lingüística ni de las generalizaciones psicológicas intencionales.¹⁹

4. La estructura semántica interna y la individuación del contenido

Si bien los enfoques presentados en las secciones previas no ofrecen teorías semánticas detalladas, comparten una tesis semántica general: el contenido de los conceptos está estructurado. Al postular una estructura conceptual interna, estos enfoques postulan que los conceptos tienen partes y que cada parte contribuye a determinar el contenido del concepto. Así, según estos enfoques el contenido de los conceptos depende de sus relaciones con otros conceptos. Tanto las teorías que proponen conceptos estructurados a partir de rasgos definicionales (condiciones necesarias y/o suficientes para caer bajo un concepto), como las que proponen rasgos no definicionales (rasgos estadísticos, como en el caso de los prototipos y rasgos de instancias particulares, como en el caso de los ejemplares) adoptan

¹⁹ Los defensores del enfoque de ejemplares no niegan que pueda haber, además de los conjuntos de ejemplares, representaciones generales de cada categoría. Dependiendo de cómo se las caracterice, estas representaciones podrían ser las encargadas de explicar la publicidad.

alguna forma de teoría semántica de rol conceptual. Estas teorías sostienen que la determinación del contenido de los conceptos primitivos está, al menos parcialmente, constituida por relaciones inferenciales, esto es, los conceptos se determinan por las inferencias en las que participan. En el caso de los conceptos entendidos como definiciones, las inferencias constitutivas son modales, en tanto cada rasgo de la definición constituye una condición necesaria y el conjunto de todos los rasgos una condición suficiente para que algo caiga bajo el concepto. En el caso de los conceptos entendidos como prototipos, las inferencias constitutivas son estadísticas: tener un concepto es conocer qué propiedades o características exhiben típicamente las cosas a las que éste se aplica (Fodor, 1999). El caso de los ejemplares es similar, pero las inferencias refieren a las instancias particulares de la categoría a partir de las cuales está estructurado cada concepto.

De este modo, los enfoques comparten la tesis de que el contenido de un concepto estará determinado por sus relaciones inferenciales con otros conceptos. Ahora bien, una objeción a este tipo de propuesta señala las dificultades para delimitar, entre todas las inferencias en las que participan los conceptos, un conjunto de relaciones constitutivas. Si no disponemos de un modo para determinarlas, entonces todas las relaciones inferenciales entre conceptos serían constitutivas de su contenido. Esto conduce al holismo, dado que el contenido de un concepto dependería de todas las relaciones que mantiene con otros conceptos. Como vimos en el capítulo 1, el holismo acarrea consecuencias indeseables, puesto que los conceptos no serían compartibles. Dado que no es probable que distintos individuos compartan sistemas inferenciales completos, puesto que éstos involucran a todas sus creencias, dos individuos no podrían tener los mismos conceptos, así como tampoco un individuo en dos momentos distintos (Fodor & Lepore 1992, Fodor 1999).

El problema de estos enfoques no sería que no pueden dar cuenta del contenido, sino que, por el modo en que lo determinan, se ven amenazados por el holismo. Y esto, a su vez, acarrea dificultades con el requisito de publicidad. Como señalé en el capítulo 1, si los conceptos no son compartibles, entonces no es posible formular generalizaciones psicológicas que dependan del contenido de los pensamientos. En particular, no sería posible formular explicaciones que apelen a los contenidos de las actitudes proposicionales (como creencias y deseos) para explicar la conducta. Asimismo, tampoco sería posible explicar la comunicación lingüística. Puesto que generalmente se asume que los significados de las emisiones lingüísticas son conceptos, si no compartimos conceptos, no compartimos los significados lingüísticos. Esto, a su vez, tornaría la comunicación lingüística una misteriosa casualidad.

Una respuesta usual a este tipo de objeciones propone que hay un modo de delimitar aquellas relaciones que son constitutivas del contenido. En este sentido, los enfoques presentados adoptan posturas localistas: sólo algunas de las relaciones de un concepto con otros resultan constitutivas de su contenido. Los modelos psicológicos de prototipos y ejemplares adoptan este tipo de propuesta, puesto que asumen que los conceptos tienen una estructura limitada de partes. Por ejemplo, en los modelos de prototipos los conceptos están estructurados como conjuntos limitado de rasgos estadísticos. Los rasgos constitutivos de un concepto son sólo aquellos que resultan salientes y tienen mayor relevancia diagnóstica. Como señala Weiskopf (2009b), el hecho de que haya modelos psicológicos que postulan una estructura semántica interna y que, como vimos en §2.1 y §3.1 resultan adecuados para dar cuenta de una serie de fenómenos psicológicos y cuentan con apoyo empírico, sugiere que una respuesta localista a los argumentos holistas es posible.

5. Las ventajas de combinar definiciones, prototipos y ejemplares con otros tipos de estructuras

En este capítulo evalué las principales ventajas y desventajas de los enfoques de prototipos y ejemplares. El objetivo de este análisis es doble: a partir de las limitaciones que presenta cada tipo de estructura para satisfacer los requisitos fundamentales de una teoría de conceptos, intento mostrar que se deben postular múltiples tipos de estructuras conceptuales y, a la vez, a partir de sus ventajas, intento defender que la clase de los conceptos incluye tanto estructuras de tipo prototípico como ejemplares, entre otras. Defenderé esto en detalle en el capítulo 7. Sin embargo, ya se pueden ir estableciendo algunas consideraciones y delimitaciones en favor del pluralismo de conceptos a partir de lo visto en este capítulo.

La evidencia presentada y analizada en este capítulo sugiere que los prototipos y los ejemplares intervienen en determinados procesos cognitivos conceptuales, especialmente en los procesos de categorización y que, dependiendo de características específicas de la tarea, podemos recurrir a uno u otro tipo de estructura conceptual. Por otra parte, las dificultades que presentan ambos enfoques sugieren que es preciso incorporar otros tipos de estructuras para dar cuenta de los requisitos explicativos fundamentales de una teoría de conceptos.

Ahora bien, con respecto a los prototipos y los conjuntos de ejemplares, cabe preguntarse, luego del análisis de sus dificultades, por qué habrían de incorporarse en una teoría estructuras conceptuales que no son capaces de satisfacer requisitos centrales como la composicionalidad. Si bien me ocuparé en el capítulo 7 de desarrollar cómo da cuenta de este requisito la propuesta pluralista que propongo, adelantaré aquí un argumento general al

respecto. Una primera consideración tiene que ver con cómo se interpreta el requisito de composicionalidad. Como señalé en el capítulo 1, puede adoptarse una lectura débil, según la cual no es necesario que en todos los casos de combinación de conceptos el contenido quede exhaustivamente determinado por los conceptos combinados más las reglas de combinación, sino que puede haber otros procedimientos de combinación de conceptos. Como señalan Prinz (2002) y Robbins (2002), puesto que la motivación para postular un requisito de composicionalidad es poder explicar la sistematicidad y la productividad del pensamiento, la combinación composicional debe ser posible. Debemos ser capaces de combinar conceptos productivamente y de interpretar un concepto compuesto recurriendo únicamente a sus constituyentes y las reglas. Pero no es necesario que esto sea así en todos los casos, sino que podemos recurrir a otras estrategias para combinar conceptos. En este sentido, se debe mostrar que la composicionalidad es posible, pero no que es el único modo de combinar conceptos.

Ahora bien, dado el análisis presentado en este capítulo, como se mencionó, ni los prototipos ni los ejemplares pueden garantizar la composicionalidad. Los modelos ofrecidos para combinar prototipos sólo consiguen dar cuenta de casos muy sencillos y no consiguen formular reglas de combinación que permitan una combinación productiva y sistemática como la que exhibe nuestra mente. La situación de los ejemplares es incluso menos promisoria. Sin embargo, esto no implica que debamos descartarlos como parte de la clase de los conceptos. Los efectos de tipicidad resultan robustos a través de categorías de distinto tipo, al igual que los efectos asociados a nuestros encuentros con instancias de las categorías, lo que sugiere que, para muchos procesos cognitivos conceptuales, recurrimos a prototipos o ejemplares. La conclusión que sugiero extraer a partir de las dificultades de estos dos enfoques con el requisito de composicionalidad es que es necesario incluir en una teoría pluralista de conceptos otros tipos de representaciones que garanticen la satisfacción de este requisito. Como defenderé en el capítulo 4, las representaciones atómicas parecen ser las indicadas para esta función.

Puede ofrecerse un argumento similar en relación a las dificultades con el requisito de alcance. La imposibilidad de representar todas las categorías por medio de prototipos o ejemplares deja de ser un problema cuando se adopta una perspectiva pluralista. El enfoque pluralista que adoptaré no sólo abandona la tesis de la uniformidad, según la cual la clase de los conceptos es homogénea, sino también la tesis de la singularidad, según la cual cada categoría está representada por un único concepto (Weiskopf, 2009a). Este enfoque propone que para cada categoría podemos tener una diversidad de estructuras conceptuales

heterogéneas. Podemos tener, por ejemplo, un prototipo de RÍO y un conjunto de ejemplares de ríos (así como representaciones de redes causales relativas a los rasgos de los ríos, de las que me ocuparé en el próximo capítulo). Pero esto no significa que para cada categoría tengamos necesariamente una estructura conceptual de cada tipo. Para algunas categorías podemos carecer de ejemplares, porque nunca encontramos ninguna de sus instancias. De este modo, las dificultades que enfrentan las teorías de prototipos y ejemplares si se las considera de manera aislada, se reducen cuando se propone complementarlas con otros tipos de representaciones, como intentaré mostrar a lo largo de esta tesis.

Con respecto a las definiciones, su *estatus* psicológico ha sido muy cuestionado. Como vimos, los efectos de tipicidad y de ejemplares parecen sugerir que no recurrimos a ellas en tareas conceptuales fundamentales como la categorización, sino que usamos para esto otros tipos de estructuras. Sin embargo, éstas podrían resultar adecuadas para algunos conceptos, como los lógicos y matemáticos. De hecho, los efectos observados en tareas de categorización no deberían ser el único *test* de estructura conceptual. Los requisitos fundamentales abarcan otras consideraciones y parece difícil dar cuenta del contenido de los conceptos lógicos y matemáticos apelando a conjuntos de ejemplares o a prototipos (así como tampoco por medio de representaciones atómicas, como sostendré en el capítulo 4). De modo que sugeriré que los conceptos también pueden tener estructura definicional, aunque ésta esté limitada a algunos dominios particulares.

En el capítulo siguiente me ocuparé del enfoque de teoría-teoría. Al igual que respecto de los enfoques de definiciones, prototipos y ejemplares, analizaré sus ventajas en relación a algunos de los requisitos fundamentales, pero mostraré que tiene dificultades con el requisito de contenido.

Capítulo 3

Tipos de estructura interna. Segunda parte: las teorías

En el capítulo 2 analicé tres propuestas de estructura conceptual interna: las definiciones, los prototipos y los conjuntos de ejemplares. Respecto de las definiciones, señalé los argumentos y la evidencia que condujeron a su abandono como una teoría general de conceptos. Sin embargo, sostuve que puede resultar adecuada para la representación de algunos conceptos, en particular los lógicos y matemáticos, de modo que se puede postular que algunos conceptos tienen estructura definicional. Con respecto a los prototipos y los ejemplares señalé, por un lado, las ventajas y la evidencia que se ha ofrecido a favor de cada tipo de estructura y, por otro lado, los inconvenientes que enfrenta cada una en relación a algunos de los requisitos. Concluí que hay evidencia sólida para incorporar ambos tipos de estructura en una teoría de conceptos y sugerí cómo se puede hacer frente a algunas de sus limitaciones al combinarlas con otros tipos de representaciones. Señalé que no es necesario que un solo tipo de estructura dé cuenta de todos los efectos observados, sino que, en cambio, distintos tipos de estructuras pueden intervenir en distintas tareas o en distintos contextos, dando lugar a perfiles de desempeño distintos. Así, puede defenderse que las definiciones, los prototipos y los ejemplares se complementan.

En este capítulo me ocuparé de otros tipos de estructuras conceptuales, las propuestas por el enfoque llamado de teoría-teoría. Este enfoque señala la insuficiencia de los enfoques de prototipos y ejemplares para dar cuenta de ciertos efectos observados en tareas de categorización en adultos, así como de determinados patrones de cambio en el desempeño en esas tareas a lo largo del desarrollo. También señala que las estructuras de rasgos meramente correlacionados por su ocurrencia conjunta no resultan suficientes y que los conceptos deben incorporar relaciones causales y, según algunas versiones, esencias ocultas de las categorías.

En §1 abordaré las tesis generales del enfoque. En este enfoque se agrupan diversas teorías heterogéneas, que pueden clasificarse en dos vertientes, una según la cual los conceptos son análogos a los términos teóricos, en tanto se individualizan por el rol que cumplen en teorías, de la cual me ocuparé en §2, y otra según la cual los conceptos en sí mismos son análogos a teorías, que analizaré en §3. En estas secciones me ocuparé de detallar las tesis específicas de cada una de las vertientes y evaluaré la evidencia a su favor. En §4 señalaré algunas dificultades que enfrenta este enfoque, en sus dos vertientes, fundamentalmente en relación a los requisitos de contenido y alcance. Por último, en §5, a la luz de la evidencia considerada, señalaré los elementos de este enfoque que considero que resulta fructífero

incorporar en una propuesta pluralista, así como el tipo de respuesta que puede ofrecer una propuesta de este tipo a las dificultades con los requisitos de contenido y alcance que enfrentan las teorías cuando se las considera de manera aislada.

1. El enfoque de teoría-teoría de los conceptos

El enfoque de la teoría-teoría surgió en parte como respuesta a los enfoques de prototipos y ejemplares, según los cuales, como vimos, por un lado, los conceptos consisten en un conjunto de rasgos relacionados por su ocurrencia conjunta y, por otro lado, las operaciones conceptuales, tanto de categorización como de combinación de conceptos, se basan en la comparación de rasgos. A diferencia de estos enfoques, que se concentran fundamentalmente en el requisito de categorización, gran parte de los autores que adoptan el enfoque de la teoría-teoría concentran su atención en el requisito de adquisición. Les interesa determinar si los niños y los adultos comparten conceptos, qué cambios conceptuales se observan a lo largo del desarrollo, así como también describir los mecanismos y procesos que dan lugar a tales cambios. Los autores agrupados en este enfoque observaron una serie de efectos en estudios relacionados tanto con el desarrollo y el cambio conceptual en los niños (Carey 1999, Carey & Spelke 1994, Gopnik & Meltzoff 1997, Keil 1989a) como con la categorización y combinación de conceptos en adultos (Ahn & Luhmann 2004, Murphy y Medin 1999, Rehder 2003, Rehder & Burnett 2005), que resultan difíciles de explicar por los enfoques basados en la similitud. Sin embargo, a pesar de coincidir en que los enfoques de prototipos y ejemplares resultan insuficientes, en general los defensores de la teoría-teoría no cuestionan la corrección de tales enfoques (e.g. Ahn 1998, Keil 1989a, Murphy & Medin 1999).

Las distintas propuestas que se agrupan en este enfoque coinciden en tres tesis centrales, relacionadas entre sí: (i) la estructura conceptual *explica* las características de los miembros de la clase representada, (ii) los distintos rasgos constitutivos de un concepto están relacionados causalmente, y (iii) los procesos conceptuales involucran razonamientos, especialmente inferencias a la mejor explicación. En relación a la primera, los defensores de este enfoque se oponen al énfasis puesto en rasgos típicos superficiales, accidentales o perceptivos y, en cambio, enfatizan la relación entre conceptos y teorías, proponiendo que las teorías proveen explicaciones para las frecuencias y correlaciones entre rasgos. Así, proponen que la estructura conceptual, en lugar de estar basada en la similitud, recoge centralmente aquellos atributos que permiten explicar el resto de las características de la categoría. Respecto a la segunda tesis, para poder explicar los atributos de la categoría, los rasgos

constitutivos de los conceptos deben estar causalmente relacionados, no pueden estar limitados a un conjunto que no especifique las relaciones entre ellos. En consecuencia, esta perspectiva postula que los conceptos forman parte de, o son en sí mismos (dependiendo, como veremos en este capítulo, de la versión del enfoque adoptada), estructuras de relaciones causales. Y, por último, los procesos que operan sobre representaciones de este tipo no pueden estar limitados al cómputo de similitudes entre rasgos, sino que son procesos de otro tipo. En particular, estos autores proponen que el procesamiento conceptual se asemeja más bien a la resolución de problemas e involucra razonamientos, en particular del tipo de las inferencias a la mejor explicación.

La caracterización de la noción de “teoría” ha sido motivo de extensos debates en filosofía de las ciencias. En términos generales, de las diversas características que se pueden atribuir a las teorías, los defensores del enfoque de la teoría-teoría destacan principalmente una: su función explicativa. Una de las caracterizaciones más detallada de las propiedades de las teorías que inspiran a este enfoque se debe a Gopnik y Meltzoff (1997). Esta caracterización constituye un buen punto de partida para abordar este enfoque, puesto que la mayoría de las propiedades que atribuyen a las teorías pueden rastrearse en las distintas propuestas que el mismo engloba. Las teorías se construyen en relación a determinada evidencia y esto, como afirman Gopnik y Meltzoff, se refleja en sus propiedades estructurales, funcionales y dinámicas. Estructuralmente, primero, las teorías son abstractas, en el sentido en que el vocabulario en que se formulan es distinto al vocabulario de la evidencia en la que se basan, y las entidades y leyes que postulan “explican los datos, pero no son meras reformulaciones de los mismos” (Gopnik & Meltzoff 1997, p. 35). Segundo, son coherentes, en tanto hay relaciones sistemáticas entre las distintas entidades postuladas por una teoría y la evidencia. Tercero, apelan a relaciones causales subyacentes que se suponen responsables de los datos observados. Cuarto, las teorías implican compromisos ontológicos, en tanto se supone que las clases que postulan cortan el mundo por sus articulaciones, i.e. son clases naturales, y soportan contrafácticos acerca de las características que tendrían y cómo se comportarían las entidades postuladas en circunstancias distintas a las reales.

Con respecto a sus características funcionales, primero, las teorías permiten hacer predicciones que van más allá de la evidencia que llevó a su construcción. Segundo, ofrecen interpretaciones de la evidencia, en lugar de meras descripciones. Tercero, ofrecen explicaciones basadas en características estructurales, i.e. explican por medio de una estructura abstracta, coherente y causal que da cuenta de un fenómeno. Por último, las teorías tienen características dinámicas, relativas a sus procesos de formación y cambio. El factor

que mayor influencia tiene en el cambio de una teoría (o en su abandono y la formación de una nueva) es la acumulación de evidencia en su contra. Sin embargo, los procesos de cambio son complejos y las teorías no se abandonan instantáneamente ante el surgimiento de evidencia en su contra. Se pueden identificar distintas etapas frente a la evidencia en contra de una teoría que incluyen, a grandes rasgos, ignorar esa evidencia, no considerarla relevante, desarrollar hipótesis *ad hoc* para lidiar con ella, complejizar la teoría inicial agregando hipótesis auxiliares que a menudo afectan su coherencia interna, desarrollar modelos alternativos para dar cuenta de la evidencia en cuestión (lo cual puede involucrar cambios conceptuales significativos), y períodos de experimentación y observación.

Algunas de estas características atraviesan a todas las versiones de teoría-teoría de los conceptos. En particular, la apelación a relaciones causales es un rasgo central que destacan los distintos autores para explicar tanto fenómenos relativos al cambio conceptual en el desarrollo, como efectos en la categorización en adultos y efectos asociados a la combinación de conceptos. Más allá del acuerdo en las tesis centrales, la teoría-teoría engloba una serie de propuestas heterogéneas, entre las cuales se pueden identificar dos vertientes principales. La primera propone que los conceptos son análogos a términos teóricos: los conceptos forman parte de teorías y se individualizan por los roles conceptuales que cumplen en las teorías en las que figuran (Carey 1985, Gopnik & Meltzoff 1997, Murphy 2000, Murphy & Medin 1999). Según esta vertiente, los conceptos “forman parte de esquemas más amplios de representación de conocimiento y están restringidos por el contenido de ese conocimiento” (Murphy 2000 p. 361). Esta vertiente enfatiza la necesidad de explicar la coherencia conceptual, i.e. de dar cuenta de por qué cuando se agrupan determinadas instancias bajo una categoría, estas agrupaciones “parecen tener sentido para quien las percibe” (Murphy y Medin 1999 p. 427). Según Murphy y Medin, una categoría es coherente cuando sus rasgos no sólo están correlacionados en el sentido de ocurrir conjuntamente con frecuencia, sino que hay razones para esa correlación, hay una explicación causal que vincula esos atributos y que forma parte de alguna teoría acerca del mundo.

La segunda, en cambio, caracteriza a los conceptos como “teorías en miniatura”, i.e. sostiene que los conceptos en sí mismos están estructurados como teorías y almacenan conocimiento de relaciones nomológicas, causales o funcionales sobre las categorías que representan (Ahn & Luhmann 2004, Keil 1989a, Rehder 2003, Rehder & Burnett 2005). Por un lado, los autores de esta vertiente señalan que hay una asimetría en la relevancia que tienen los distintos rasgos constitutivos de un concepto para determinar la pertenencia a una categoría. Esta asimetría no depende de la correlación entre rasgos ni de su utilidad

diagnóstica (tal como ocurre en las teorías de prototipos), sino de su centralidad. Por ejemplo, el hecho de tener una cierta forma curva resulta altamente relevante para identificar un boomerang, mientras que no es así para identificar una banana, a pesar de que la curvatura es un rasgo igualmente típico de los boomerangs y las bananas (Medin & Shoben 1988). Los defensores de esta vertiente coinciden en atribuir a las representaciones conceptuales una estructura interna que relaciona los rasgos causalmente, en la cual la centralidad de un rasgo depende de las relaciones de causa-efecto que mantiene con otros rasgos (Ahn 1998, Ahn & Luhmann 2004, Keil 1986, 1989a, Medin & Shoben 1988, Rehder 2003). Algunos de ellos consideran que además de la estructura causal, los conceptos incorporan representaciones acerca de propiedades “ocultas” de las categorías. Distintos estudios indican que atribuimos a los miembros de las categorías ciertas propiedades ocultas o esencias que permanecen estables aunque se modifiquen las propiedades superficiales. Por ejemplo, si se modifica la apariencia de un mapache para que parezca un zorrillo, seguimos considerando que es un mapache en virtud de ciertas propiedades ocultas (aunque no sepamos exactamente en qué consisten). A esta versión se la denomina “esencialismo psicológico” (Gelman, 2004; Keil, 1986).

Es importante hacer algunas aclaraciones sobre la distinción entre las dos vertientes. Los autores a menudo oscilan entre una y otra posición, o caracterizan a los conceptos en términos de una u otra, para luego adoptar una posición más bien intermedia. Así, por ejemplo, Murphy y Medin (1999) sostienen que para caracterizar los conceptos es preciso atender a sus roles en teorías, pero también aluden a una estructura interna de rasgos causalmente relacionados. De manera inversa, Keil (1989) afirma que los conceptos son en sí mismos teorías, pero también apela a sus vínculos con teorías más amplias para explicar ciertos efectos. A pesar de estas oscilaciones, la distinción entre vertientes cumple una función a la hora de presentar las distintas teorías, puesto que los autores de una y otra apelan a distintos niveles explicativos para dar cuenta de las características de los conceptos. Si bien ninguno niega que, por un lado, los conceptos pueden tener una estructura interna y, por otro lado, los conceptos participan en la construcción de teorías, hay diferencias en la relevancia explicativa que asignan a uno u otro aspecto. Una aclaración de Carey respecto de en qué casos tiene sentido hablar de distintos conceptos en los niños y los adultos, resulta ilustrativa de en qué nivel individúa los conceptos la primera vertiente:

[...] tener distintos conceptos debe distinguirse de tener distintas creencias.

Nadie dudaría que los niños tienen distintas creencias que los adultos. [...] Los

conceptos pueden diferir a lo largo de múltiples dimensiones; no hay duda de que no hay un único sentido de ‘diferentes conceptos’ que defender. Asumo que hay un continuo de grados de diferencia conceptual, en un extremo del cual están los conceptos embebidos en sistemas conceptuales inconmensurables. (1988, p. 168)

En este pasaje se refleja en qué consiste la adhesión a la vertiente de los conceptos como partes de teorías. Carey, al igual que Murphy y Medin o Gopnik y Meltzoff, evalúa las diferencias entre conceptos en función del rol que cumplen en esquemas más amplios de conocimiento. Los defensores de la vertiente de conceptos como teorías, en cambio, individualizan los conceptos en función de su estructura interna. Como veremos en las próximas secciones, los autores de una y otra vertiente explican los efectos relevantes en términos distintos, apelando, los primeros, a las teorías generales y, los segundos, a la estructura conceptual interna.

Por otra parte, como mencioné anteriormente, los autores que trabajan en el enfoque de teoría-teoría no niegan que los conceptos puedan tener otros tipos de estructura, e.g. prototípica, pero sí señalan las limitaciones de los conjuntos de rasgos estadísticamente correlacionados para dar cuenta de la evidencia que repasaré en las próximas secciones. En este sentido, algunas de las propuestas podrían interpretarse como pluralistas. Sin embargo, como se verá en este capítulo, los defensores de este enfoque se han concentrado en establecer el rol que cumplen las teorías en la adquisición y uso de los conceptos, y no han desarrollado articulaciones de este enfoque con otros, como el de los prototipos. Si bien en muchos de los trabajos se aclara que no pretenden negar que los conceptos puedan tener, por ejemplo, prototipos, la mayoría de las teorías que se han desarrollado en este enfoque no se proponen como extensiones o complementos de los enfoques previos, sino como alternativas.²⁰

2. Los conceptos como partes de teorías

En esta sección reconstruiré la primera vertiente de la teoría-teoría, según la cual los conceptos satisfacen sus distintas funciones en virtud de sus vínculos con estructuras más amplias de conocimiento de tipo teórico. Se han propuesto distintas fuentes de evidencia a favor de esta vertiente, que abarcan tanto estudios del desarrollo cognitivo en los niños, como

²⁰ Una excepción importante es la teoría de Keil (1989a; 1989b), quien, como veremos en §3, sugiere una articulación entre la representación de correlaciones estadísticas y la estructura teórica.

del desempeño en adultos. En §2.1 reconstruiré las tesis principales de esta vertiente y en §2.2 mencionaré las principales fuentes de evidencia a favor de la misma.

2.1 Los vínculos entre los conceptos y las estructuras más amplias de conocimiento

La tesis central de esta vertiente afirma que los conceptos forman parte de estructuras teóricas de conocimiento, y que su adquisición, desarrollo y aplicación, depende de la adquisición, desarrollo y características de las teorías más amplias a las que pertenecen. En una lectura débil, el hecho de que los conceptos forman parte de teorías no parece una tesis que los distintos autores que postulan representaciones mentales estén dispuestos a discutir. Carey, por ejemplo, caracteriza las relaciones entre conceptos, creencias y teorías de la siguiente manera:

Por *concepto*, *creencia* y *teoría*, me refiero a aspectos de representaciones mentales. Los conceptos son unidades de representación mental, a grandes rasgos del grano de ítems léxicos simples, tales como *objeto*, *materia*, *peso*. Las creencias son proposiciones representadas mentalmente que el creyente toma como verdaderas, tales como *El aire no está hecho de materia*. Los conceptos son los constituyentes de las creencias; esto es, las proposiciones están representadas por estructuras de conceptos. Las teorías son estructuras mentales complejas que consisten en un dominio de fenómenos mentalmente representados y principios explicativos que dan cuenta de ellos. (1992, p. 89)

Esta caracterización parece un punto de partida común a todos los enfoques representacionistas, independientemente de los debates epistemológicos acerca de cómo están estructuradas las teorías. Lo que diferencia a la primera vertiente de la teoría-teoría de otros enfoques representacionistas de conceptos es el rol que atribuyen a las teorías en la individuación, adquisición, desarrollo y uso de los conceptos. En este sentido, la primera vertiente se ha desarrollado en dos versiones distintas, una más fuerte y una más débil, dependiendo de cómo se individúan los conceptos (Weiskopf 2009c). En su versión fuerte, esta tesis afirma que el contenido de los conceptos está determinado por sus roles en las teorías de las que forman parte. Por ejemplo, Carey (1985, p. 198), afirma que “los conceptos

deben identificarse por los roles que cumplen en teorías”.²¹ De la misma manera, Murphy y Medin (1999) señalan que entender los conceptos como equivalentes a la suma de sus componentes resulta problemático y que, en cambio, debe pensarse en las representaciones conceptuales como estructuras embebidas en conocimiento teórico acerca del mundo. El rol que tienen las teorías en la individuación de conceptos en esta lectura no se limita a restringir su contenido, sino que resultan constitutivas del mismo:

El enfoque basado en teorías expande los límites de las representaciones conceptuales: para caracterizar el conocimiento y el uso de un concepto debemos incluir todas las relaciones en las que ese concepto está involucrado y el resto de los conceptos que dependen de él. Para explicar la coherencia conceptual, deben considerarse los procesos que operan sobre un concepto, además de la información directamente almacenada en el mismo. (Murphy & Medin, 1999, p. 436)

En su versión débil, en cambio, esta vertiente de la teoría-teoría afirmaría solamente que tenemos cuerpos de conocimiento proposicional teórico almacenado y que los conceptos forman parte de ese conocimiento (Weiskopf, 2009c). Esta última lectura parece casi trivial y afirma una tesis a la que todos los autores que postulan representaciones mentales adhieren. El punto es si el rol que cumplen los conceptos en las teorías influye de manera significativa en su aplicación en tareas como la categorización o la combinación conceptual, i.e. si las estructuras teóricas de las que los conceptos forman parte determinan el uso que hacemos de los conceptos y son, en este sentido, constitutivas de los mismos. Considero que una lectura más adecuada de la versión débil de esta vertiente es la que ofrece Machery (2009), quien señala que la misma se compromete con la tesis según la cual los conceptos están organizados por dominios, i.e. que les atribuimos determinadas características en función de su pertenencia a clases más amplias, como la de los seres vivos o los artefactos. El contenido de los conceptos estaría así restringido o guiado por las propiedades relativas a las teorías acerca del dominio al que pertenecen, pero no completamente determinado por ellas. A

²¹ Si bien su postura se ha ido modificando relativamente a lo largo del tiempo, en sus trabajos más recientes, Carey (2009) continúa defendiendo que el contenido de los conceptos depende de su rol en esquemas de inferencia conceptual, aunque ha debilitado esta tesis adoptando una teoría del contenido de doble factor, que sostiene que el contenido de los conceptos está determinado en parte por las entidades a las que el concepto refiere y en parte por su rol conceptual.

menudo los autores oscilan entre la interpretación fuerte y débil de esta tesis, de modo que en lo que sigue no retomaré la distinción.

El vínculo entre conceptos y teorías tal como se propone en esta vertiente encuentra dos motivaciones principales. Por un lado, la observación de cambios conceptuales a lo largo del desarrollo llevó a algunos autores a proponer que el desarrollo conceptual en los niños es similar al cambio teórico científico (Carey 1985, Gopnik & Meltzoff 1997). Por otro lado, consideraciones sobre el rol de la coherencia entre atributos en la delimitación de categorías, así como en la combinación de conceptos, llevó a postular que las características que atribuimos a las categorías dependen de cuerpos de conocimiento teórico de trasfondo (Murphy 2000, Murphy & Medin 1999). A continuación abordaré cada una de estas motivaciones.

En relación al desarrollo conceptual y el cambio teórico, Gopnik y Meltzoff (1997) sostienen que su apelación a teorías para caracterizar los conceptos no pretende ser meramente metafórica, sino que los mismos mecanismos cognitivos intervienen en el desarrollo científico y el desarrollo conceptual en los niños. Esta postura sugiere que evolutivamente venimos equipados con ciertas habilidades cognitivas, de las cuales un conjunto especial está destinado a la explicación, inferencia, predicción y atribución causal con el fin de comprender el mundo que nos rodea. En términos de los autores, estas habilidades nos permiten descubrir y describir verdades (Gopnik & Meltzoff 1997). Estas habilidades se ponen en práctica desde una edad muy temprana, los niños muy pequeños tienen representaciones abstractas del mundo y teorías intuitivas, que van cambiando a lo largo del desarrollo en función de la experiencia y el aprendizaje. Desde esta perspectiva, los niños no son pequeños científicos, sino que los científicos son grandes niños “[...] tomando mecanismos naturales de cambio conceptual, diseñados para facilitar el aprendizaje en la infancia, y poniéndolos en práctica en un modo culturalmente organizado.” (Gopnik & Meltzoff 1997, p. 21).

Según las primeras propuestas de la teoría-teoría, los procesos de aprendizaje y desarrollo conceptual consistían fundamentalmente en la construcción y puesta a prueba de hipótesis (Gopnik & Meltzoff, 1997). Esta propuesta enfrentaba un problema de pobreza del estímulo, puesto que la evidencia disponible difícilmente resulta suficiente para brindar apoyo a una hipótesis y descartar otras, sino que es compatible con múltiples hipótesis mutuamente inconsistentes. Para responder a estos problemas, la estrategia usual fue la postulación de una base teórica innata lo suficientemente rica como para restringir la

formulación de las primeras hipótesis.²² Por ejemplo, para el concepto OBJETO, se propuso que la percepción de objetos está guiada desde el nacimiento por ciertos principios acerca de los objetos físicos, en particular por los principios de continuidad espaciotemporal y solidez. Estos principios proporcionan la estructura teórica para el concepto OBJETO, del cual serían propiedades nucleares innatas. Esta propuesta postula principios innatos para determinados dominios, a los que caracteriza como *conocimiento nuclear* y sostiene que a lo largo del desarrollo los conceptos se van enriqueciendo con otros principios, para cuya adquisición pueden formar parte los procesos de construcción de teorías, pero que las representaciones conceptuales no se adquieren exclusivamente a partir de la observación de evidencia ni todas se modifican a la manera de las teorías científicas (Carey 2009, Carey & Spelke 1996, Spelke, Breinlinger, Macomber & Jacobson 1992, Spelke & Kinzler 2007).

Recientemente, ha recibido considerable desarrollo una respuesta alternativa que apela al aprendizaje probabilístico y propone que los niños forman representaciones estructuradas como modelos causales, en los cuales no sólo se representan un conjunto de rasgos y los mecanismos causales que relacionan esos rasgos, sino también las relaciones entre la estructura causal y la evidencia (Gopnik 2011, Gopnik & Schulz 2004, Gopnik & Wellman 2012, Kushnir, Gopnik, Lucas & Schulz 2010, Lucas, Bridgers, Griffiths & Gopnik 2014). Estas propuestas se concentran en explicar cómo la evidencia estadística, así como la intervención y la experimentación, permiten construir modelos causales.

En relación a la segunda motivación de esta vertiente de la teoría-teoría, i.e. dar cuenta de ciertos efectos de coherencia conceptual, Murphy y Medin (1999) señalan que para explicar las características de los conceptos es preciso apelar a teorías en dos niveles. Por un lado, un concepto será coherente, estará integrado o “tendrá sentido”, en la medida en que los distintos rasgos que constituyen su estructura interna estén conectados en términos de causas y efectos, así como de estructuras y funciones. Por otro lado, la coherencia también está determinada por el rol que cumple un concepto en las estructuras más amplias de conocimiento de trasfondo. Sin estos roles, los conceptos se tornan inestables o resultan olvidados. Así, según Murphy y Medin (1999), cada concepto estará determinado por el poder explicativo que tenga la agrupación de determinados atributos en relación al mundo. En consonancia con esta tesis, la categorización de un objeto bajo una categoría no depende de la medida en que ambos comparten una serie de atributos, sino de la medida en que los atributos de la categoría guardan una relación explicativa con los atributos del objeto. Por ejemplo, si

²² Una compilación de propuestas de este tipo se encuentra en Hirschfeld y Gelman (1994).

vemos que, en una fiesta, una persona se tira a una pileta vestida, probablemente la categoricemos como una persona borracha, no porque sus atributos sean similares a los atributos de otras personas borrachas, o a los comportamientos típicos de las personas borrachas, sino porque la borrachera nos permite explicar su comportamiento.

2.2 Evidencia a favor de los vínculos constitutivos entre conceptos y teorías

Una fuente de evidencia a favor de esta vertiente proviene de estudios del desarrollo. Carey (1985) presenta un extenso estudio de la adquisición y el desarrollo de los conceptos relativos al dominio de los seres vivos, por ejemplo ANIMAL, PLANTA, PERSONA, VIVO, MUERTE y CRECIMIENTO. Tomando como ejemplo el caso de ANIMAL, Carey señala que antes de los 5 años de edad, los niños centran su concepto de animal en la capacidad de actuar, i.e. en el hecho de que su comportamiento es generado por sí mismo, en oposición a los seres inanimados. Carey caracteriza la teoría de los niños como una *psicología ingenua*, en tanto explica los comportamientos biológicos como respirar o digerir en términos de necesidades, deseos y planes (Carey 1985), o como una *biología vitalista*, en tanto la propiedad central que atribuye a los animales es la de llevar a cabo actividades que están generadas internamente (Carey 1988). Este tipo de teoría se diferencia de la *biología mecanicista* que tenemos desde aproximadamente los 10 años, en la medida en que esta última incluye nociones como la necesidad de reproducción y alimento, así como detalles sobre cómo se resuelven estas necesidades en cada especie y nociones mecanicistas respecto de distintos sistemas (respiratorio, digestivo, circulatorio) como sistemas corporales interrelacionados responsables del mantenimiento del organismo, de las cuales las primeras teorías carecen.

Entre los 4 y los 10 años, habría una serie de cambios en la teoría sobre la biología. Por ejemplo, Carey (1988) señala que antes de los 5 años, los conceptos de VIDA y MUERTE están indiferenciados de los de ANIMADO e INANIMADO, respectivamente. Para esto, ofrece como evidencia ejemplos de razonamientos que llevan a cabo los niños de esa edad. Por ejemplo, su hija de 3 años y medio considera que su osa de peluche no tiene sangre ni huesos, que no puede morir y siempre va a estar viva. Sin embargo, cuando le preguntan si está viva, inmediatamente dice que no, que está muerta, pero expresa confusión al respecto. Finalmente, concluye que como tiene un tamaño mediano, está entre viva y muerta, y a veces se mueve. En otros ejemplos, su hija pregunta cómo hace la gente muerta para ir al baño o comenta que es extraño que las estatuas no están vivas y sin embargo se las pueda ver (Carey 1988). Carey señala que a esa edad los niños no diferencian no-vivo,

muerto e inanimado y que, en este sentido, sus conceptos de seres vivos son inconmensurables con los de los adultos, en el sentido kuhniano en que dos teorías científicas pueden ser inconmensurables.

Otras fuentes de evidencia a favor de esta vertiente se construyen por oposición a los enfoques de prototipos y ejemplares, en tanto intentan registrar efectos que no se pueden explicar apelando a la coocurrencia de rasgos en el mundo ni al cómputo de similitudes entre representaciones. Por ejemplo, Murphy (2000, 2002) reporta evidencia de estudios que indican que resulta más fácil aprender categorías nuevas cuando sus atributos están relacionados y forman un conjunto coherente, que cuando sus rasgos son inconsistentes o las relaciones entre ellos resultan neutrales. Por ejemplo, una categoría cuyos atributos son estar hecho en África, tener un aislamiento leve, ser verde, circular por junglas y tener ruedas, resulta más fácil de aprender que una cuyos atributos son ser verde, tener transmisión manual, tener cubiertas radiales, tener *airbags* y tener asientos de vinilo. Los dos conjuntos de rasgos corresponden a vehículos y ninguno presenta características inconsistentes, pero los rasgos del primer conjunto tienen entre sí una relación temática específica: remiten a un vehículo de la jungla. Murphy señala que lo que facilita el aprendizaje de las categorías cuyos rasgos están temáticamente relacionados es el hecho de poder establecer relaciones explicativas específicas entre los rasgos, puesto que cada rasgo puede vincularse a una estructura común de conocimiento y esto ayuda a recordar cómo están agrupados los rasgos. Así, cuanto más fuerte o marcada resulte la coherencia entre los rasgos, i.e. cuanta mayor cantidad de vínculos se puedan establecer entre los rasgos y una estructura causal subyacente a ellos, más fácil resulta el aprendizaje de una categoría. “Cuando la gente estudia miembros de una categoría, busca explicaciones de por qué los ítems son como son”, afirma Murphy (2000, p. 372).

Con respecto al rol de la coherencia en la estructura conceptual, se pueden mencionar también estudios relativos a la combinación de conceptos. Murphy y Medin (1999) sostienen que el conocimiento de trasfondo acerca de cómo funciona el mundo juega un rol fundamental a la hora de combinar conceptos. Para referirse a los mecanismos que operan en la combinación de conceptos, los autores apelan a la metáfora de las reacciones químicas, en lugar de apelar a la adición de componentes. Citando uno de los ejemplos que ofrecen, “la reparación de un experto” es, probablemente, una reparación llevada a cabo por un experto, mientras que “la reparación de un motor” no es una reparación llevada a cabo por un motor. La combinación de dos conceptos no se limita a la intersección de los rasgos comunes a ambos, ni a la suma de los rasgos de cada uno, sino que parece haber una construcción basada en conocimiento más general. Según Murphy y Medin, la interpretación de un concepto

compuesto involucra formular hipótesis a partir de teorías generales, respecto de qué aporta cada concepto al compuesto.

Como vimos en el capítulo 2, un fenómeno que ha resultado difícil de acomodar por parte de los enfoques de prototipos es el de las propiedades emergentes. Un ejemplo clásico en la literatura, es el de los **CARPINTEROS EGRESADOS DE HARVARD**, que comenté brevemente en el capítulo 1. Se atribuye a los carpinteros egresados de Harvard el no ser materialistas, a pesar de que aquella no es una característica atribuida generalmente a los egresados de Harvard ni a los carpinteros. En este caso, el no ser materialistas es una propiedad emergente (Kunda, Miller & Claire 1990). Kunda y sus colegas llevaron a cabo un estudio concentrándose en la combinación de conceptos que representan categorías sociales y, fundamentalmente, en casos en los que parece haber conflicto entre las representaciones asociadas a una y otra categoría, como el caso de los carpinteros y los egresados de Harvard.

Defendiendo una tesis similar a la de Murphy y Medin, Kunda y sus colegas proponen que las propiedades emergentes se explican porque en la combinación de conceptos intervienen razonamientos causales y formulaciones de hipótesis explicativas. Según Kunda y sus colegas, cuando se combinan dos conceptos, la determinación de los atributos de las instancias que caen bajo el concepto compuesto implica, por lo menos en algunos casos, la construcción de una explicación causal de las razones por las cuales esas instancias pertenecen a las dos categorías (i.e., a cada una de las categorías que se combinan). Las propiedades emergentes (o la pérdida de propiedades de los conceptos iniciales) son el resultado de estos razonamientos causales y en ellos intervienen tanto relaciones entre los atributos de cada concepto como razonamientos generales que involucran el conocimiento de teorías más amplias.²³

A continuación abordaré la otra vertiente de la teoría-teoría. Como veremos en 3.2, el tipo de evidencia que se ha propuesto a favor de esa vertiente resulta en muchos casos similar al revisado en esta sección.

²³ Los ejemplos que proponen Kunda y sus colegas corresponden estrictamente a categorías sociales y, vale decir, resultan en algunos casos incómodos por su explícita apelación a prejuicios sociales. La lista de ejemplos incluye, entre otros, “maratonista ciego”, “abogado ciego” y “hombre de negocios de izquierda”, como ejemplos de conceptos que “sorprenden” por las características que se atribuyen a cada uno de los conceptos combinados y requieren, en consecuencia, de un razonamiento que justifique la inclusión de una instancia bajo tal concepto. De todos modos, incluyo esta evidencia porque la tesis de que la combinación de conceptos involucra razonamientos y explicaciones causales acerca de la pertenencia de una instancia a más de una categoría resulta útil para explicar efectos observados en la combinación de conceptos en general, en tanto el fenómeno de las propiedades emergentes no se limita a los tipos de casos presentados en este estudio.

3. Los conceptos como teorías

En lo que sigue reconstruiré la segunda vertiente de la teoría-teoría, según la cual los conceptos en sí mismos están estructurados como teorías. De manera análoga a la presentación en 2, en 2.1 presentaré las tesis principales de las distintas propuestas que se agrupan bajo esta vertiente y en 2.2 los principales estudios que dan apoyo a la misma.

3.1 Estructuras de rasgos relacionados causalmente

En términos generales, esta vertiente sostiene las mismas tesis principales que la anterior, mencionadas en 2: la estructura conceptual está constituida por los rasgos que explican los atributos de la categoría representada, estos rasgos están causalmente relacionados y los procesos conceptuales se asemejan a la formulación de inferencias a la mejor explicación. Sin embargo, esta vertiente interpreta estas tesis en términos de una estructura conceptual interna de rasgos causalmente relacionados, en lugar de enfocarse en los vínculos entre conceptos y estructuras más amplias de conocimiento. En lo que sigue desarrollaré con mayor detalle cada una de estas tesis a partir de la propuesta de Keil, uno de los representantes principales de esta vertiente.

Al igual que la mayoría de las posturas abordadas en las secciones anteriores, Keil (1983; 1986; 1989; 1991) desarrolla su propuesta por oposición a las teorías que atribuyen a los conceptos exclusivamente una estructura de propiedades típicas. Se opone también a las versiones de psicología del desarrollo que postulan un cambio a partir de una estructura basada exclusivamente en la similitud y la coocurrencia de rasgos a una estructura teórica, y propone en cambio que hay estructuras teóricas que restringen el contenido de los conceptos desde la adquisición inicial de conceptos. Según Keil los conceptos incluyen rasgos y relaciones que “van más allá” de las características superficiales.

Keil (1991) sostiene que no todos los conceptos exhiben una estructura causal compleja, sino que se puede establecer un continuo desde conceptos estructurados en una serie de rasgos que no están conectados causalmente hasta aquellos que se estructuran como conjuntos complejos de propiedades causalmente relacionadas, que pueden caracterizarse en términos de cúmulos de propiedades homeostáticas que son causalmente responsables de las propiedades superficiales de los miembros de la clase. Estos últimos serían los conceptos de clases naturales. Aquí Keil adopta la propuesta de Boyd (1991) según la cual las clases naturales, en sentido ontológico, son aquellas cuyos miembros comparten un cúmulo homeostático de propiedades, i.e. una serie de propiedades más o menos estables, que permiten establecer generalizaciones acerca de todos los miembros. La clave para que estas

propiedades permitan delimitar una clase parece radicar en su rol causal: estas propiedades serían causalmente responsables del resto de las propiedades de los miembros de la clase.

Pues bien, Keil propone incorporar esta idea para la estructura psicológica de los conceptos y sostiene que nuestros conceptos de clases naturales están estructurados en función de mecanismos causales que permiten explicar las propiedades de la clase. Aquí se ven reflejadas dos de las tesis características de la teoría-teoría, mencionadas en 1: por un lado, la tesis de que la estructura de los conceptos es aquella que permite explicar las características de los miembros de la clase representada y, por otro lado, la idea de que, para explicar tales características, los mecanismos causales resultan especialmente relevantes. Keil (1989b, p. 47) afirma que “la mayoría de los conceptos son teorías parciales en sí mismos, en tanto incorporan explicaciones de las relaciones entre sus constituyentes, de sus orígenes y de sus relaciones con otros cúmulos de propiedades”.

Según Keil (1991), no todos los conceptos presentan el mismo tipo de estructura, sino que habría un continuo desde los conceptos de clases nominales, pasando por los de clases de artefactos simples y artefactos complejos, hasta los de clases naturales. Keil caracteriza como clases nominales a aquellas que están delimitadas por una construcción social o convencional, como los conceptos TÍO o TRIÁNGULO, a diferencia de las clases naturales, que recogen regularidades del mundo natural. Las clases nominales, afirma, no tienen estructuras causales complejas que vinculen una serie de características centrales o críticas y una serie de características superficiales.

A continuación, detallaré las características específicas que los principales representantes de esta vertiente atribuyen a las representaciones conceptuales en función de la evidencia que ofrece cada uno.

3.2 Evidencia a favor de la estructura interna análoga a estructuras teóricas

Numerosos estudios aportan evidencia a favor de distintas tesis de esta vertiente. Uno de los primeros estudios evalúa patrones casos de categorización en adultos. Rips (1989) llevó a cabo una serie de estudios con el objeto de mostrar una doble disociación entre los juicios de categorización y los de similitud. En los estudios a los sujetos se les pedía que evaluaran o bien la posible pertenencia de una instancia a dos categorías, o bien de cuál de las dos categorías resultaba más típica, o bien su similitud respecto de las mismas. Por ejemplo, uno de los casos presentaba como categorías alternativas a las pizzas y las monedas de 25¢. La primera tarea de los participantes era estimar el diámetro de la pizza más pequeña y de la moneda de 25¢ más grande que recordaran. A partir de estas respuestas, los investigadores

presentaban luego un caso de un objeto con la misma forma, de un diámetro que era exactamente la mitad entre los valores estimados en la primera tarea, y presentaban una de las tres tareas, i.e. determinar si es una pizza o una moneda, respecto de cuál de las dos categorías resulta más típico, o si es más similar a una pizza o a una moneda.

En el diseño del experimento siempre se presentaba una categoría para la cual la dimensión en cuestión suele ser variable y otra para la cual suele ser fija. En el caso pizza-moneda, el diámetro de las pizzas suele ser variable, mientras que el diámetro de las monedas de 25¢ suele estar fijado por leyes o regulaciones estatales. Los casos presentados incluían, entre otros, los pares de categorías tetetas/pelotitas de tenis, y cantidad de integrantes de una banda de rock/cantidad de integrantes del senado de EEUU. Con respecto a la pertenencia a las categorías, en la mayoría de los casos la respuesta consideraba más probable que el objeto en cuestión perteneciera a la categoría de dimensión variable, mientras que, inversamente, con respecto a la similitud, la mayoría de las respuestas favorecía la categoría de dimensión fija. Por último, con respecto a la tipicidad, las respuestas resultaron prácticamente azarosas, i.e. aproximadamente la mitad de las respuestas consideraba al objeto más típico de la categoría de dimensión variable y la otra mitad lo consideraba típico de la categoría de dimensión fija.

Esta evidencia de doble disociación entre los juicios de categorización y de similitud apoya la tesis de que la categorización no consiste en el cómputo de similitudes, sino que debe incluir otro tipo de proceso. Rips sostiene que categorizar se asemeja a formular inferencias a la mejor explicación y, para defender esta tesis, apela a las justificaciones de los sujetos del estudio respecto de sus juicios de categorización. Para citar un ejemplo, en un caso en que se debía evaluar si un objeto de 145cm es una caja de cereal o una señal de tránsito de detención, la respuesta de uno de los participantes es:

Probablemente sea una caja de cereal enorme, porque nadie vería una señal de detenerse si midiera tan poco... Una señal de detenerse debería tener una cierta altura, y si bien es improbable encontrar una caja de cereal tan grande, no tendría sentido tener una señal de tránsito tan pequeña. (Rips, 1989, p. 37)

Rips señala que la categorización, en este caso, no se limita a mencionar como una propiedad “esencial” de las señales de tránsito el tener determinada altura, sino que involucra un razonamiento relativo a la función que cumple la altura para satisfacer los propósitos de las señales de tránsito. En este sentido, aunque la categoría de las señales de tránsito incluya

como una propiedad central el tener una determinada altura, la evaluación de si algo cae o no bajo esta categoría no se limita a comparar alturas, sino que implica evaluar si el objeto en cuestión puede satisfacer las funciones de las señales de tránsito y esto involucra la construcción de una inferencia explicativa.

Otra fuente de evidencia proviene de estudios de psicología del desarrollo, que sugieren la relevancia de rasgos distintos de los perceptivos en la categorización. Especialmente para el dominio de los seres vivos, Keil (1986; 1989) señala que desde muy temprana edad tenemos una tendencia a basar los juicios de categorización en algo más que las características superficiales y típicas de los miembros de una categoría, ya sea en una estructura causal subyacente o una esencia de la categoría. Keil muestra, en numerosos estudios, evidencia a favor de tal estructura causal, así como un cambio de énfasis a lo largo del desarrollo cognitivo desde la categorización a partir de las características típicas superficiales en los primeros años, a la asignación progresiva de mayor relevancia a los principios causales y las esencias ocultas.

Por ejemplo, en uno de los estudios más conocidos al respecto, se planteaba a niños de distintas edades una situación en la que un grupo de doctores había realizado ciertas operaciones para cambiar la apariencia de determinados individuos. Por ejemplo, habían modificado un caballo para que pareciera una cebra (pintándolo como una cebra, cortándole la crin, etc.), y lo habían entrenado luego para que se comportara como una cebra (para que comiera pasto en lugar de avena, etc.). Los niños debían determinar a qué categoría pertenecía esa instancia luego de la transformación. Estos estudios fueron realizados con ejemplos de clases naturales y artificiales.²⁴ Una de las predicciones de la teoría de Keil era que, para las clases naturales, los niños más pequeños (alrededor de 5 años) juzgarían que las instancias cambiaron de categoría, a diferencia de los niños más grandes (entre 7 y 10 años), que no juzgarían el cambio de categoría por el cambio de apariencia, mientras que, para las clases artificiales, todos juzgarían el cambio de categoría. Los resultados confirmaron estas predicciones (Keil 1989a). Estos estudios sugieren, por un lado, una diferencia en la estructura de rasgos para las clases biológicas y los artefactos, según la cual los conceptos de clases biológicas incluyen de manera central determinados rasgos ocultos que resultan relevantes para determinar la pertenencia a la categoría.

²⁴ En esta serie de estudios, las modificaciones involucraban siempre el cambio de apariencia de un animal a otro animal (como el caso de caballo/cebra) y de un artefacto a otro artefacto (por ejemplo tacho de basura/silla), no se producían pasajes de clase natural a artificial ni viceversa.

Los resultados de estos estudios podrían interpretarse a favor de la versión según la cual los niños comienzan estructurando los conceptos exclusivamente a partir de las características superficiales de sus miembros. Sin embargo, otra serie de estudios con escenarios de transformación sugiere que a pesar de que los niños pequeños están dispuestos a aceptar la transformación de una especie en otra, la respuesta es negativa si la transformación atraviesa, por ejemplo, categorías de animales a plantas o de animales de juguete a animales vivos. Si bien los niños pequeños conceden la transformación entre especies, aun entre especies sumamente diferentes, como mariposas y peces, no aceptan que un puercoespín pueda convertirse en un cactus, a pesar de las similitudes superficiales entre ambos. En este sentido, algunos principios teóricos respecto de categorías más amplias, como las de seres vivos, artefactos, animales y plantas parecen estar en juego (Keil, 1989a). La tesis de Keil es que los patrones de cambio en las respuestas en función de la edad, en lugar de explicarse por un cambio cualitativo que consistiría en un pasaje de conceptos estructurados a partir de la similitud de rasgos superficiales a conceptos estructurados como teorías, se explican en cambio en virtud del enriquecimiento de las teorías disponibles, sin apelar a un estado inicial en el que los conceptos están estructurados exclusivamente en función de rasgos superficiales.

Rips (1989) llevó a cabo estudios con escenarios similares de transformación de clases naturales y artificiales en adultos, en los que los participantes debían responder no sólo a qué categoría pertenecía una instancia luego de la transformación, sino también cuán similar era a cada una de las categorías y cuán típica de cada una de ellas (utilizando el mismo tipo de consigna que en los estudios reconstruidos más arriba). Como resultado, observó una disociación entre las evaluaciones de similitud, tipicidad y pertenencia a la categoría, apoyando la tesis de que la evaluación de similitud no es el proceso central que guía los juicios de categorización.

En otra serie de estudios relacionados a las diferencias de los conceptos por dominios, se reportaron distintos patrones de respuesta en tareas de verificación de propiedades para categorías de seres vivos y categorías de artefactos. Por ejemplo, en una serie de estudios, Keil (1983) observó que cuando a niños pequeños se les indicaba que los “hyraxes” estaban somnolientos, los niños, a pesar de no saber qué eran los hyraxes, afirmaban que los hyraxes también debían e.g. poder tener hambre, y no podían estar enchufados o rotos. Este tipo de resultados se ha interpretado a menudo como evidencia de la influencia de conocimiento de dominio específico en la atribución de propiedades a las categorías. A partir de estos y otros resultados, algunos autores de este enfoque infieren una estructura distinta para los conceptos

de seres vivos y los de artefactos, y sostienen que distintos tipos de rasgos resultan más centrales en distintos tipos de categorías (Keil, 1989a; Rips, 1989). Asimismo señalan los tipos de rasgos cualitativamente distintos que resultan centrales para las clases naturales y los artefactos. Así, afirman que para las clases naturales resultan más centrales los rasgos de composición o moleculares, i.e. los rasgos relativos a la composición estructural interna, mientras que para las clases de artefactos los rasgos más centrales son los funcionales, i.e. los rasgos relativos a las actividades del artefacto y a los propósitos e intenciones de los diseñadores.

Otros autores, en cambio, sostienen que tales efectos de interacción entre tipos de rasgos y dominios se explican por el *estatus causal* de los rasgos (Ahn 1998, Ahn, Gelman, Amsterlaw, Hohenstein & Kalish 2000, Ahn & Luhmann 2004). La hipótesis del estatus causal sostiene que, independientemente de si los rasgos son de composición o son funcionales, así como del dominio en cuestión, lo que determina la centralidad de los rasgos es su relevancia causal. Según esta hipótesis un rasgo será más central en la medida en que su instanciación cause la instanciación de otros rasgos y, a la inversa, será menos central en la medida en que su instanciación dependa de que se instancien otros rasgos que lo causan. Ahn y sus colegas llevaron a cabo una serie de estudios para poner a prueba esta hipótesis, en los cuales se crearon categorías novedosas, que correspondían a clases naturales o a artefactos. Cada objeto se describía en términos de rasgos composicionales (su estructura molecular, su estructura de partes internas, su sustancia o su origen) y funcionales (aquello que hace o para qué se usa) y se manipularon los estímulos para que los rasgos composicionales causaran rasgos funcionales o viceversa.

Por ejemplo, una de las categorías creada para el estudio es la de una especie de hormiga, la Kehoe. Se indica a los individuos que el 80% de las hormigas de esta especie tiene un alto nivel de sulfato de hierro en la sangre y que el 80% de las hormigas Kehoe digiere la comida rápido. Luego se les presenta o bien una relación en la que el rasgo de composición es causa del funcional, como “el nivel alto de sulfato de hierro en sangre tiende a causar una rápida digestión de comida; el sulfato de hierro en la sangre de las hormigas Kehoe estimula las enzimas responsables de la producción de las secreciones para digerir comida y una hormiga Kehoe puede digerir la comida más rápido con más secreciones” (Ahn 1998, p. 171); o bien una relación en la que el rasgo funcional es causa del de composición, como “la digestión rápida de comida en las hormigas Kehoe tiende a causar un alto nivel de sulfato de hierro en la sangre; la digestión de comida acelera el metabolismo de las hormigas Kehoe lo que resulta en el alto nivel de sulfato de hierro en la sangre” (Ahn 1998, p. 171). A

continuación, se presentaban dos preguntas, cuán probable será que una hormiga que tiene sangre con un alto nivel de sulfato de hierro pero no digiere rápido sea una hormiga Kehoe y, a la inversa, cuán probable será que una hormiga que digiere la comida rápido, pero no tiene un nivel alto de sulfato de hierro en la sangre, sea una hormiga Kehoe.

De modo análogo, para las categorías de artefactos, se presentaba un tipo de sistema séptico llamado Kehoe, que tiene líquido rico en sulfato de hierro y descompone los residuos cloacales rápido. Así, manipulando el tipo de rasgo, categoría y su estatus causal en ejemplos con categorías diseñadas para los estudios, Ahn y sus colegas reportaron resultados consistentes con la hipótesis del status causal (Ahn 1998, Ahn *et al.* 2000, Ahn & Luhmann 2004). Según estos resultados, los participantes juzgan más probable la pertenencia de una instancia a una categoría si comparte los rasgos que son causalmente centrales que si comparte los rasgos que son efectos, más allá de si los rasgos son de composición o funcionales y de si la categoría en cuestión es una clase natural (aunque ficticia, diseñada para el experimento) o un artefacto.

Otro efecto de la estructura causal reportado en la literatura es el de coherencia. Este efecto consiste en una tendencia a considerar más probable que una instancia pertenezca a una categoría si presenta ciertas combinaciones de rasgos coherentes con la estructura causal de la categoría. En este sentido, un ejemplar que presenta rasgos que representan causas sin sus respectivos efectos (o viceversa, rasgos que son efectos sin sus respectivas causas) es considerado menos representativo de una categoría que uno que presenta una combinación de causas y efectos consistentes con las relaciones causales de la categoría. Así, la relevancia de cada rasgo para determinar la pertenencia a una categoría no depende sólo de su status causal, sino también de su coherencia con el resto de los rasgos (Rehder 2003, Rehder & Kim 2006, 2010).

Se han propuesto distintos modelos para dar cuenta de estos efectos, así como de su interacción.²⁵ Más allá del debate entre los distintos modelos de este enfoque, los resultados sintetizados en esta sección evidencian, por un lado, las limitaciones de las estructuras basadas exclusivamente en la similitud y en la coocurrencia de rasgos (como las de prototipos o ejemplares), y, por otro lado, sugieren que las relaciones causales entre rasgos resultan cruciales para explicar el desempeño en determinadas tareas, así como los patrones de desarrollo conceptual.

²⁵ Para estudios de la interacción entre los efectos de status causal y de coherencia, así como un modelo que permite dar cuenta de ellos, véase Hayes y Rehder (2012).

En síntesis, la evidencia repasada brevemente en esta sección y en §2.2, sugiere que las estructuras de relaciones causales ocupan un lugar significativo en nuestro sistema conceptual. Ya sea que se adopte la primera o la segunda vertiente, y más allá de las diferencias particulares entre los distintos modelos propuestos, los estudios mencionados brindan un apoyo considerable a las tesis centrales de este enfoque. Sin embargo, en lo que sigue señalaré ciertas dificultades que enfrenta, fundamentalmente, en relación al requisito de contenido.

4. Dificultades del enfoque de teoría-teoría

Las teorías que se agrupan en el enfoque de teoría-teoría, en cualquiera de sus vertientes, tienen dificultades para dar cuenta del requisito de contenido. Este es un requisito del cual generalmente este enfoque no se ocupa explícitamente. Sin embargo, cualquier teoría que asuma que los conceptos son representaciones mentales de categorías, como es el caso de las que se agrupan en la perspectiva de la teoría-teoría, acarrea compromisos semánticos, de modo que resulta pertinente explicitar estos compromisos y, partiendo de ellos, evaluar si pueden dar cuenta de este requisito. En lo que sigue desarrollaré tres problemas que enfrentan las vertientes de la teoría-teoría en relación a la determinación del contenido de los conceptos.

Primero, ambas vertientes de la teoría-teoría implican un compromiso, por lo menos parcial, con una semántica de rol conceptual, según la cual el contenido está determinado por los roles inferenciales de los conceptos, i.e. por las inferencias en las que estos participan. Esto es así porque ambas vertientes implican que la individuación del contenido de un concepto dependerá de sus relaciones con otros conceptos. En el caso de la primera vertiente, el contenido de un concepto dependerá de las relaciones con otros conceptos que están implicadas por la teoría a la que el concepto pertenece, mientras que, en la segunda vertiente, dependerá de las relaciones con los rasgos a partir de los cuales está estructurado el concepto, i.e. de las relaciones con los rasgos constitutivos del concepto. Esto resulta problemático, puesto que el sistema conceptual se vería amenazado por el holismo y, en consecuencia, los conceptos serían inestables.

Como vimos en el capítulo 2, la semántica de rol conceptual o inferencial puede conducir al holismo, ante la falta de un método que permita distinguir las inferencias que resultan relevantes para individuar el contenido de un concepto de aquellas que no, y, en consecuencia, *todas* las inferencias contribuirán a individuar el contenido (Fodor 1987, Fodor & Lepore 1992). Como vimos en las secciones previas, algunos autores de este enfoque se

comprometen explícitamente con el holismo. Este es el caso, por ejemplo, de Murphy y Medin (1999), quienes afirman que los límites de los conceptos deben extenderse hasta incorporar *todas* las inferencias en las que estos participan. Un compromiso de este tipo acarrea un sistema holista que, a su vez, conduce a la inestabilidad de los conceptos, puesto que cualquier cambio en una creencia constituirá un cambio conceptual. Como consecuencia, por un lado, los conceptos no serán compartidos por distintos individuos, puesto que es poco probable que compartamos sistemas completos de creencias. Por otro lado, los conceptos no serán estables en la vida mental de un individuo. En términos de Fodor (1999, p. 221) “... la semántica de rol inferencial hace intratable el problema de la individuación de conceptos”.

Una posibilidad sería intentar resolver estas dificultades proponiendo una interpretación localista de las relaciones constitutivas de los conceptos, i.e. sostener que sólo algunas de las relaciones con otros conceptos resultan constitutivas del contenido. En este sentido, una tarea de la psicología cognitiva podría ser la de determinar un subconjunto de relaciones conceptuales estables que pueda considerarse constitutivo del contenido (Weiskopf 2009b). Ahora bien, aun si la propuesta localista funcionara, pudiendo evitar el holismo y la consecuente inestabilidad conceptual, la teoría-teoría enfrenta otros dos problemas en relación a la determinación del contenido.

El segundo problema radica en que para muchos conceptos no disponemos de teorías completas y detalladas a la manera de las teorías científicas, sino que disponemos de teorías más bien esquemáticas e incompletas. De acuerdo a la teoría-teoría, en muchos casos la pertenencia o no de un individuo a una categoría depende de si posee cierta esencia oculta. Sin embargo, si no tenemos representaciones específicas de, por ejemplo, la esencia de PERRO y de GATO (más allá de alguna representación de MATERIAL GENÉTICO), entonces no queda claro qué determina que una teoría sea acerca de perros o acerca de gatos (Margolis & Laurence 2003). Por esta razón, debe haber un modo de individuar el contenido de los conceptos más allá de su estructura teórica. Las estructuras teóricas (en cualquiera de sus vertientes) no resultan suficientes para fijar la referencia de los conceptos. Para esto es necesario recurrir a otro tipo de representaciones. El tercer problema también está relacionado con la identificación de las esencias. En muchos casos las teorías especifican las esencias de manera circular, por ejemplo, al identificar los tigres como aquellos que tienen esencia de tigre y son crías de tigres (Prinz, 2002). De esta manera, un concepto estructurado en función de la esencia oculta de los tigres tendrá información acerca de tigres sólo si se dispone de otra representación que permita fijar la referencia de TIGRE.

Como consecuencia de estas dificultades, aunque las teorías pueden codificar información acerca de la pertenencia a categorías, adoptar únicamente conceptos estructurados como teorías o conceptos individuados únicamente por su rol en teorías (dependiendo de la vertiente), no permitiría dar cuenta del requisito de contenido. Así, para dar cuenta de este requisito, el enfoque de la teoría-teoría debe suplementarse con representaciones estructuradas de otro modo, o representaciones atómicas, sin estructura semántica interna, que permitan fijar la referencia de las teorías. Una propuesta híbrida o pluralista de conceptos que incorpore conceptos estructurados como teorías cuenta con una ventaja para resolver estas dificultades, dado que dispone de otros tipos de representaciones, como se verá en la próxima sección.

Este enfoque también presenta dificultades en relación al requisito de alcance. Como señala Weiskopf (2009c), tenemos teorías articuladas para muy pocos conceptos. Probablemente, tengamos conocimiento teórico acerca de electrones, hipopótamos y océanos, pero éste no parece ser el caso para conceptos como VERDAD. Sin embargo, los defensores de este enfoque disponen de respuestas posibles a esta objeción. Estas consisten, por un lado, en restringir el alcance de la teoría-teoría a algunos tipos de conceptos, admitiendo que otros conceptos pueden tener estructuras de otro tipo y, por otro lado, en interpretar “teoría” de un modo lo suficientemente amplio como para poder abarcar las estructuras de aquellos conceptos que no exhiben una estructura teórica completamente desarrollada. Representantes de ambas vertientes han sostenido alguna de estas dos alternativas. Algunos de los defensores de la primera vertiente restringen el alcance de los vínculos constitutivos entre conceptos y teorías a algunos dominios, como el de la biología, la física y la psicología, y admiten que disponemos de otros tipos de estructuras conceptuales, por ejemplo prototipos (Carey 1985, Carey & Spelke 1996). Otros autores que defienden esta misma vertiente aclaran que usan “teoría para referirse a cualquiera de una colección de ‘explicaciones’ mentales, en lugar de una versión científica completa y organizada. [...] el término connota un conjunto complejo de relaciones entre conceptos, usualmente con una base causal” (Murphy & Medin 1999, p. 426).

De manera análoga, los defensores de la vertiente de los conceptos como teorías, consideran que no todos los conceptos involucran conocimiento teórico detallado, sino que reconocen que ciertos conceptos se organizan en torno a principios causales y otros, por ejemplo, en torno a convenciones sociales. Keil, por ejemplo, como vimos en §3, propone un continuo de conceptos, en el cual las estructuras teóricas tienen distinto grado de relevancia y complejidad. Por otro lado, adoptando la segunda estrategia, se puede señalar que, aun en el

extremo menos teórico del continuo, la estructura conceptual involucra relaciones entre los rasgos y, en este sentido, si bien algunos conceptos no están constituidos por teorías completamente articuladas y detalladas, serían teorías en este sentido mínimo. Así, las dificultades con respecto al requisito de alcance no parecen tan significativas como las relativas al requisito de contenido.

5. La integración de teorías en una propuesta pluralista

Como vimos, el espectro de propuestas incluido en la teoría-teoría es sumamente heterogéneo. Parece adecuado afirmar que todas las propuestas comparten una serie de intuiciones cruciales respecto a la estructura conceptual: ésta no se limita a una serie de rasgos superficiales que ocurren conjuntamente, sino que debe incorporar relaciones causales que permitan explicar las características de los miembros de la categoría que representa. Asimismo, los procesos conceptuales involucran una serie de habilidades cognitivas similares a las que se ponen en juego en el desarrollo de teorías científicas, específicamente, los procesos de adquisición de conceptos involucran la formulación (y posiblemente reformulación a lo largo del desarrollo) de teorías, y categorizar equivale a realizar inferencias inductivas, en algunos casos inferencias a la mejor explicación.

La dificultad principal que enfrenta este enfoque se relaciona con la determinación del contenido de los conceptos. Dados los tres problemas relativos a la individuación del contenido de las representaciones teóricas, analizados en el apartado anterior, si se pretende incorporar teorías como estructuras semánticas en una teoría de tipo pluralista, es preciso, por un lado, limitar el conjunto de rasgos o inferencias que resultan constitutivos del contenido conceptual y, por otro lado, incorporar representaciones distintas de las teorías, que permitan fijar la referencia. Las teorías pluralistas que incorporan estructuras de tipo teórico han propuesto un modo de fijar la referencia que consiste en complementar la propuesta de estructura conceptual con una semántica que permita individuar el contenido de los conceptos de modo independiente de las relaciones que mantienen con otros conceptos (Margolis & Laurence 2003, Prinz 2002, Weiskopf 2009b).

La clave de estas propuestas consiste en postular que la estructura interna no es la que determina la referencia de los conceptos, sino que apelan para esto a relaciones de covariación como las que propone la semántica informacional, de la que me ocuparé en detalle en el próximo capítulo. Según este tipo de semántica el contenido está constituido por relaciones nomológicas entre la mente y el mundo. Así, el contenido de un concepto estaría determinado por la propiedad en el mundo con la que covaría, y no por sus relaciones con

otros conceptos (Fodor 1999). Ahora bien, esta solución postula dos tipos de contenido conceptual: uno que fija la referencia en función de relaciones causales o nomológicas con el mundo, al que Weiskopf (2009b) llama “contenido referencial”, y otro que depende de la estructura interna del concepto (y, por lo tanto, de sus relaciones con otros conceptos), al que llama “contenido cognitivo”. El contenido referencial sería aquello que los conceptos representan, mientras que el contenido cognitivo sería cómo lo representan y estaría estructurado en un conjunto de rasgos (que pueden tener entre sí distintas relaciones de dependencia, incluyendo relaciones causales, distinciones entre rasgos centrales y periféricos o entre esencias y rasgos superficiales, como sugiere la teoría -teoría). Como veremos en el capítulo 5, Prinz (2002) propone una distinción similar entre dos tipos de contenido, el contenido real, que refiere a las esencias de las categorías, y el contenido nominal, que refiere a las propiedades a partir de las cuales identificamos las instancias que caen bajo un concepto, y puede estar estructurado a la manera de las teorías.

Este tipo de propuesta parece adecuada, puesto que permitiría incorporar representaciones de tipo teórico y así, dar cuenta de la evidencia de los diversos efectos observados en las tareas de categorización y los cambios a lo largo del desarrollo y, a la vez, podría eludir los tres problemas señalados. Sin embargo, como desarrollaré en el capítulo 6, en relación a la teoría de Prinz, este tipo de propuesta tiene sus propias dificultades. Por un lado, el problema del holismo continúa aquejando al contenido cognitivo. Una propuesta de este tipo aún debe resolver cómo acotar el conjunto de rasgos o inferencias que resultan constitutivos del contenido cognitivo de los conceptos. En relación a los otros dos problemas, i.e. al hecho de que muchas teorías son esquemáticas y algunas caracterizan las esencias de modo circular, sí dispone de una respuesta satisfactoria, puesto que la fijación de la referencia de los conceptos no dependerá de cuán detalladas o específicas son las teorías, sino de relaciones nomológicas con el mundo. Por otra parte, esta propuesta enfrenta un nuevo problema que consiste en explicar la articulación entre los dos tipos de contenido. Ofreceré una respuesta a este problema en el capítulo 7.

En el próximo capítulo abordaré el atomismo conceptual fodoriano que, a diferencia de todas las teorías analizadas en este capítulo y el anterior, postula que los conceptos no tienen estructura semántica interna. Como argumentaré, esta propuesta presenta numerosas ventajas, especialmente en relación a los requisitos de contenido y composicionalidad y, como anticipé en esta última sección, ofrece una teoría semántica que ha sido adoptada por algunos autores de otros enfoques. Sin embargo, como mostraré, tiene serias limitaciones para dar cuenta del cuerpo de evidencia que presenté a propósito de los enfoques de

prototipos, ejemplares y teorías, del cual una teoría adecuada de conceptos debería poder dar cuenta.

Capítulo 4

El atomismo conceptual

En los capítulos 2 y 3 reconstruí y evalué los principales tipos de estructura conceptual interna que se han propuesto: las definiciones, los prototipos, los conjuntos de ejemplares y las teorías. En este capítulo abordaré el atomismo conceptual, que rechaza la tesis de que los conceptos presentan una estructura interna y sostiene, en cambio, que la mayoría de los conceptos léxicos son atómicos (Fodor, 1999). En §1 desarrollaré las tesis principales de esta teoría, y en §2 y §3 evaluaré la satisfacción de los requisitos fundamentales. De manera análoga a los dos capítulos anteriores, mi objetivo será, por un lado, destacar las ventajas del atomismo informacional, en especial su capacidad para dar cuenta de los requisitos de alcance, contenido y composicionalidad (§2), y, por otro lado, mostrar que recurriendo únicamente a representaciones atómicas no es posible dar cuenta de los requisitos fundamentales. El atomismo conceptual tiene inconvenientes respecto del contenido en relación a la individuación de conceptos coextensivos, así como para dar cuenta de los requisitos de adquisición y categorización (§3).

En los capítulos dedicados a definiciones, prototipos, ejemplares y teorías, no me ocupé del debate en torno a los vehículos conceptuales, puesto que esos enfoques no se comprometen en general con ninguna tesis específica relativa a los vehículos. El atomismo, en cambio, incluye una tesis sobre los vehículos, según la cual estos son símbolos de un lenguaje del pensamiento. Por este motivo, en este capítulo no sólo evaluaré las tesis relacionadas con el contenido conceptual, sino también las tesis relativas al vehículo. Dedicaré especial atención a la reconstrucción de la teoría semántica que desarrolla Fodor, la semántica informacional, puesto que las teorías pluralistas suelen adoptar alguna versión de esta teoría y es la que adoptaré en la propuesta de pluralismo amplio que presentaré en el capítulo 7.

1. Las tesis principales del atomismo informacional

La propuesta del atomismo conceptual puede sintetizarse en tres tesis principales, las cuales desarrollaré en orden luego de su presentación: (i) *Lenguaje del pensamiento*: los conceptos son representaciones mentales que tienen una semántica composicional y una

sintaxis combinatoria;²⁶ (ii) *Semántica informacional*: el contenido de los conceptos se determina por medio de relaciones nomológicas con propiedades en el mundo y (iii) *Atomismo*: la mayoría de los conceptos léxicos son atómicos.

La hipótesis del lenguaje del pensamiento (LDP) o mentalés forma parte de la Teoría Representacional/Computacional de la Mente (TRCM) a la que Fodor adscribe, según la cual el pensamiento está vehiculado por un sistema interno de representaciones sobre las cuales operan causalmente los procesos computacionales. Me ocuparé de la TRCM en §3, en relación a las dificultades que tiene esta teoría para dar cuenta del requisito de categorización. A los fines de comprender la hipótesis del LDP basta con decir que, según la TRCM, los estados mentales de actitud proposicional (tales como deseos y creencias, y en general, pensamientos) consisten en relaciones (computacionales) entre organismos y representaciones mentales. Por ejemplo, mi creencia de que es invierno implica que estoy en un tipo particular de relación o actitud (en este caso, la de creencia) con una representación mental que tiene por contenido que *es invierno*.

Según la hipótesis del LDP, los contenidos de las actitudes proposicionales están conformados por un sistema interno de representaciones que es análogo a los lenguajes naturales. Se trata de un lenguaje compuesto por símbolos que se combinan para formar representaciones complejas según las reglas de combinación sintáctica y de composicionalidad semántica que rigen para los lenguajes naturales, aunque el mentalés no es ninguno de ellos (Fodor 1975, Fodor 1987). Este lenguaje, al igual que los lenguajes naturales admite, por un lado, una distinción entre sujeto y predicado y, por otro lado, ciertos operadores lógicos, como los cuantificadores y las variables. Pues bien, los conceptos son los símbolos de este LDP. Así, al igual que los enfoques evaluados en los capítulos anteriores, el atomismo informacional sostiene que los conceptos son representaciones mentales, i.e. que son particulares mentales físicamente instanciados (i.e. que tienen una realización física a nivel cerebral). Además, sostiene que éstas poseen una estructura simbólica semejante a la de un lenguaje natural. Estas representaciones poseen propiedades sintácticas y semánticas. El aspecto sintáctico de los conceptos consiste en la forma, vehículo o formato que transporta la información semántica, mientras que el aspecto semántico consiste en el contenido de los conceptos, i.e. aquello en el mundo (estados, eventos o propiedades) que los conceptos representan.

²⁶ La hipótesis del lenguaje del pensamiento está formulada para los constituyentes de los pensamientos, i.e. los conceptos. Sin embargo, cabe aclarar que si bien los conceptos son representaciones mentales, no toda representación mental es conceptual. Fodor (2008) sostiene que hay representaciones pre-conceptuales o, en terminología más habitual, no conceptuales.

Las propiedades sintácticas son propiedades que tienen los conceptos en virtud de su estructura interna. Fodor (1987, p. 18) afirma que “la sintaxis de un símbolo es una de sus propiedades físicas de orden superior. Como aproximación metafórica, podemos pensar en la estructura sintáctica de un símbolo como un rasgo abstracto de su forma.” Fodor (1987) presenta las propiedades sintácticas por analogía a las formas de las llaves y las cerraduras. Así como la forma geométrica de una llave determina qué cerradura podrá abrir, la estructura sintáctica de un concepto determina sus poderes causales, puesto que determina qué transiciones o transformaciones se podrán realizar sobre este símbolo. Por ejemplo, sobre un símbolo con la estructura P&Q puede operar una transformación que resulte en Q puramente en virtud de su forma sintáctica. Apelando a otra metáfora, Margolis y Laurence (2007) señalan que las diferencias sintácticas pueden entenderse como diferencias en la *ortografía mental*. Así, AGUA y H₂O, si bien poseen el mismo significado, dado que denotan la misma propiedad, tendrán poderes causales distintos en virtud de las diferencias en su ortografía/sintaxis.²⁷ En función de su estructura sintáctica, se puede distinguir entre representaciones atómicas y moleculares. Las representaciones moleculares son las que tienen una estructura de *constituyentes* sintácticos, que pueden ser a su vez atómicos o moleculares (Fodor & Pylyshyn, 1995). Así VIENTO, por ejemplo, es una representación atómica, mientras que LAS OLAS Y EL VIENTO es una representación molecular, que tiene como constituyentes atómicos a LAS, OLAS, Y, EL y VIENTO.

Las propiedades semánticas, a su vez, se relacionan con el contenido de los conceptos, consisten en *acerca de* qué son esas representaciones. Según la TRCM, los estados mentales de actitud proposicional tienen propiedades semánticas o condiciones de satisfacción, en función de su relación con el mundo. Por ejemplo, si creo que la Tierra es plana, el contenido de mi creencia (i.e., *la Tierra es plana*) representa un hecho o estado del mundo y es éste el que permite evaluar si mi creencia es verdadera o falsa. Dado que el contenido de estos estados mentales se compone a partir de la combinación de los contenidos de sus constituyentes (en este caso, TIERRA, ES y PLANA) los portadores primitivos del contenido intencional son los conceptos.

Para dar cuenta de cómo los conceptos adquieren las propiedades semánticas, el atomismo informacional adopta la semántica informacional. Fodor sostiene, siguiendo a Dretske (1981), que el contenido de los conceptos consiste en información acerca de propiedades en el mundo. Una representación R porta información acerca de P si las

²⁷ La sintaxis entendida en términos ortográficos es una manera de individuar los símbolos del LDP. Para otras maneras véase Schneider (2009, 2011).

instancias de *P* son “nómicamente responsables” por las instancias de *R* (Fodor 1990). Así, los conceptos adquieren su contenido en virtud de la existencia de leyes psicofísicas causales (sustentadas por contrafácticos) que relacionan instancias de nuestras representaciones mentales con instancias de propiedades en el mundo. Por ejemplo, las instancias del concepto LIEBRE caen bajo una ley según la cual las instancias de este concepto son causadas por (o serían efectos de) la propiedad de *ser liebre* instanciada en el mundo. Esta ley está sustentada por contrafácticos del tipo: Si se instanciara la propiedad de *ser liebre*, se instanciaría el concepto LIEBRE. Estas relaciones nómicas son las “hacedoras de significado”. De esta manera, Fodor adhiere a una teoría causal de la representación (aunque no apela a cadenas causales actuales, sino contrafácticas), según la cual las condiciones para la posesión de conceptos no son epistémicas, sino puramente nomológicas: no se trata de “saber algo”, sino de estar enlazado nómicamente con la propiedad adecuada. En consecuencia, tener un concepto resulta semejante a tener una etiqueta que está relacionada con una propiedad determinada. Estas etiquetas nos permiten pensar acerca de cosas, relaciones, etc.

Por otra parte, Fodor (1999) introduce una distinción entre distintos tipos de propiedades con las cuales nuestras mentes se enlazan: las *propiedades de clase natural*, por un lado, y las *propiedades de apariencia*, por otro lado. Las propiedades de clase natural son propiedades esenciales, intrínsecas, mientras que las de apariencia son propiedades que las cosas tienen en virtud de los estados mentales que evocan en mentes como las nuestras. Por ejemplo, en el caso de AGUA, la posesión de este concepto implica estar enlazado con la propiedad esencial de la sustancia agua, i.e. H_2O , por lo cual *ser agua* es una propiedad de clase natural. En cambio, en el caso de PICAPORTE, por ejemplo, no hay propiamente leyes acerca de picaportes *qua* picaportes, sino que PICAPORTE cae bajo leyes acerca de nuestra mente, puesto que *ser picaporte* es una propiedad de apariencia. No hay ninguna esencia oculta, ni propiedad intrínseca de los picaportes (tal como una microestructura) en virtud de la cual puedan caer bajo una ley que conecte a los picaportes, *qua* picaportes, con el concepto PICAPORTE, sino que *ser picaporte* es una propiedad que se define a partir de cómo ciertas cosas en el mundo resuenan en nuestras mentes, i.e. a partir de los estados mentales que despiertan en nosotros. Tanto en el caso de los conceptos de clase natural como en el de los conceptos que refieren a propiedades de apariencia, su contenido se explica por relaciones nomológicas con propiedades. De esta manera, Fodor hace extensiva la semántica informacional a la mayoría de los conceptos léxicos (ya sea que representen propiedades de clase natural o de apariencia), excluyendo de esta explicación únicamente a los conceptos lógico-matemáticos (Fodor 1999).

Se ha objetado a la semántica informacional que las relaciones causales entre el mundo y los conceptos no son suficientes para determinar el contenido, dado que no permiten casos de representación errónea. Puede ocurrir que un concepto se dispare frente a cosas que no pertenecen a su extensión. Existe la posibilidad de que, por ejemplo, se dispare el concepto MARIPOSA frente a una polilla, y según una teoría semántica puramente causal no se podría describir este caso como uno de representación errónea, puesto que habría que decir que lo que sea que causa la instanciación de MARIPOSA, pertenece a la extensión de MARIPOSA y así no habría error posible. Esto, además, da lugar a otro problema relacionado, el llamado *problema de la disyunción* (Fodor 1990, Laurence & Margolis 1999, Prinz 2002): si un concepto representa aquello que causa su instanciación y este concepto puede tener “causas fortuitas”, entonces siguiendo la semántica informacional este concepto estaría nómicamente relacionado con más de una propiedad y, en consecuencia, su contenido sería disyuntivo. En el caso del ejemplo en el cual una polilla causa MARIPOSA, el contenido de MARIPOSA sería *polilla o mariposa*. La solución que ofrece Fodor (1990) a este problema consiste en postular una dependencia asimétrica entre las relaciones nómicas “hacedoras de significado”.

Esta propuesta especifica dos condiciones que un símbolo X debe cumplir para que tenga a x como contenido. Primero, establece que es una ley que instancias de x causan instancias de X. Segundo, para todo y (distinto de x), si y *qua* y causa de hecho instancias de X, esto depende asimétricamente de que instancias de x causen instancias de X.²⁸ La primera afirmación prácticamente enuncia la tesis central de la semántica informacional. Según esta condición, hay una ley –e.g. ‘Los topos causan TOPOS’ o ‘Las vizcachas causan VIZCACHAS’– que está sustentada por contrafácticos del tipo ‘Si se estuviera en presencia de topos, estos causarían pensamientos acerca de topos’. Sin embargo, esto no resulta suficiente para resolver el problema de la disyunción. De acuerdo con la condición 1, X porta información acerca de x , pero no se asegura que porte información sólo acerca de x . Para incorporar tal restricción se agrega la segunda condición, que especifica la asimetría. Esta condición establece que las instancias de y no podrían causar instancias de X si las instancias de x no causarían instancias de X.

En otras palabras, no podríamos pensar en instancias de y como X si no pudiéramos pensar en instancias de x como X (Fodor 2008). Si en determinadas circunstancias, por

²⁸ Fodor (1990) agrega una tercera condición según la cual debe haber, de hecho, una relación causal entre algunos x y X, pero luego abandona esta condición para adoptar una semántica informacional puramente contrafáctica (Fodor, 1994).

ejemplo en una noche oscura, alguien piensa TOPO frente a un castor, la relación *castor-TOPO* depende asimétricamente de la ley que conecta las instancias de TOPO con las instancias de la propiedad *ser topo*. De esta manera puede hacerse frente al problema de la disyunción, puesto que, por la conjunción de las dos condiciones, todas las conexiones que dependen asimétricamente de otras para causar instancias de X quedan excluidas del contenido de X. De la misma manera, se resuelve también el problema del error, dado que la dependencia asimétrica restringe el conjunto de relaciones causales que determinan el contenido de las representaciones. Así, una representación de TOPO causada por un castor, porta efectivamente información acerca del castor que causó su instanciación, pero, sin embargo, no significa *castor* y es, en este sentido, una representación errónea de castores. En palabras de Fodor (1990, p. 93) “[...] hay mucho menos significado que información. Esto es así porque todo lo que se necesita para tener información es covariación causal confiable, mientras que para el significado se necesita además (por lo menos) la dependencia asimétrica.”

La tesis (iii) concierne a la estructura conceptual. Contra lo que sostienen la mayoría de las teorías de conceptos, en particular las de definiciones, prototipos, ejemplares y teorías, evaluadas en los capítulos 2 y 3, Fodor afirma que la gran mayoría de los conceptos son atómicos, i.e. no tienen estructura interna. Para la mayoría de los conceptos léxicos, la relación nómica que determina su contenido no depende de manera constitutiva de ninguna otra relación que mantengan con otros conceptos y, por lo tanto, el contenido de estos conceptos es atómico (Fodor, 1999, pp. 32-33). Para poseer, por ejemplo, el concepto GATO, no es necesario estar en relación con ninguna propiedad distinta de la de *ser gato*. Para entrar en relación (o “enlazarse”) con la propiedad de *ser gato*, las mentes como las nuestras necesitan estar en un entorno determinado, i.e un entorno en el que esté la propiedad de *ser gato*. Pero, en un entorno así, también se está en contacto con otras propiedades relacionadas con la de *ser gato*, como *ser bigote*, de modo que puede ocurrir que junto con GATO se adquieran también otros conceptos relacionados. Sin embargo, resulta contingente que se pueda adquirir GATO al mismo tiempo que BIGOTE, OREJA o PATA, de manera que estos conceptos no son constitutivos de GATO. Así, una persona podría tener el concepto GATO sin tener, por ejemplo, el concepto BIGOTE.

Fodor ha formulado numerosas críticas a las propuestas que asignan una estructura semántica interna a los conceptos, dado que, como vimos en los capítulos 2 y 3, implican adoptar alguna forma de teoría semántica de rol conceptual (Connolly *et al.* 2007, Fodor 1999, Fodor & Lepore 1996). Una semántica de rol conceptual puede conducir al holismo, lo

cual acarrea serias dificultades para la individuación de conceptos. Por otra parte, la postulación de una estructura interna también genera determinadas dificultades relativas al requisito de composicionalidad, dado que, como vimos en el capítulo 2, en muchos casos los conceptos compuestos no heredan los rasgos constitutivos de los conceptos que se combinan. Dadas estas dificultades, Fodor rechaza todos los tipos de estructura interna propuestos y defiende, en cambio, la propuesta informacional atomista, según la cual los conceptos son símbolos de un LDP que representan propiedades en el mundo en virtud de relaciones nómicas entre esas propiedades y los conceptos, y ninguna relación con otros conceptos resulta constitutiva de su contenido.

La propuesta informacional atomista ofrece significativas ventajas en relación a los requisitos de alcance, composicionalidad y contenido (aunque, al no incorporar estructura interna, tiene dificultades para dar cuenta de la individuación de conceptos coextensivos). Por otra parte, la falta de estructura interna acarrea problemas para explicar la categorización. Señalaré también que la explicación ofrecida de la adquisición resulta insuficiente. Abordaré primero las ventajas y luego las dificultades de esta propuesta.

2. Las ventajas del atomismo

En las siguientes secciones sostendré que el atomismo ofrece ventajas por su explicación del alcance, la composicionalidad y la intencionalidad.

2.1 Alcance

Una de las ventajas de esta teoría es que puede abarcar el espectro de conceptos que nuestras mentes exhiben, de modo que puede dar cuenta del requisito de alcance. Como mencioné en §1, Fodor no propone la misma teoría semántica para todos los tipos de conceptos, sino que establece una distinción entre conceptos, como AGUA, que refieren a propiedades de clase natural, conceptos como MESA que refieren a propiedades de apariencia, y conceptos como CONJUNCIÓN que son conceptos lógicos. Así, a partir de las consideraciones que Fodor presenta en distintos trabajos se puede elaborar una taxonomía, distinguiendo tres tipos de conceptos: conceptos que refieren a propiedades de clase natural, conceptos que refieren a propiedades de apariencia, y conceptos lógicos y matemáticos (1990, 1994, 1999).

Los conceptos de clase natural refieren a propiedades en el mundo en virtud de leyes psicofísicas que conectan a los conceptos con propiedades esenciales. Así, por ejemplo AGUA refiere a la propiedad H_2O , y el concepto teórico ELECTRÓN refiere a la propiedad

de *ser partícula subatómica con carga negativa*. Los conceptos que refieren a propiedades de apariencia, por otra parte, lo hacen en virtud de leyes psicofísicas causales acerca de cómo esas propiedades resuenan en nuestras mentes. Por ejemplo, CAFETERA no es un concepto de clase natural puesto que no hay una microestructura o esencia oculta de las cafeteras, pero todas las cafeteras tienen en común la propiedad de ser causas de CAFETERA en mentes como las nuestras.

De esta manera, el atomismo informacional explica satisfactoriamente el contenido de la gran mayoría de los conceptos y no tiene mayores problemas para dar cuenta del contenido de conceptos abstractos, tales como VIRTUD (a diferencia de otras teorías como la de proxitipos, que se verá en el próximo capítulo). El contenido de este concepto estará determinado por una relación nómica entre el concepto y la propiedad que causaría su instanciación (Fodor, 1990). Así, hay una propiedad abstracta que, al instanciarse bajo ciertas condiciones físicas, causa la activación de VIRTUD en mentes como las nuestras. Y, puesto que *virtud* debe ser una propiedad de apariencia (aunque Fodor no lo señale explícitamente), ésta quedaría definida como aquella que tiene la característica de causar en nuestras mentes el concepto VIRTUD. Así, tampoco es necesario en el marco de esta teoría especificar cómo es que se instancia en el mundo esta propiedad abstracta y difícil de caracterizar. No importa que distintos mecanismos puedan instanciar la virtud (podría ser una serie de acciones de cierto tipo, por ejemplo) y tampoco es necesario precisarlos. En esta explicación lo importantes es que, si se instanciara la propiedad de la *virtud*, esto causaría la activación de VIRTUD en una mente como las nuestras.

En relación a los conceptos lógicos y matemáticos, Fodor no se pronuncia explícitamente a favor de ninguna teoría semántica particular, pero sostiene que, por ejemplo, en el caso de CONJUNCIÓN, este concepto está definido por su tabla de verdad (Fodor 2008). Asumo que, a tono con el espíritu de este comentario, la manera adecuada de dar cuenta de este tipo de conceptos en la teoría fodoriana sería por medio de definiciones, puesto que para estos conceptos es posible especificar el conjunto de condiciones necesarias y suficientes para su posesión y aplicación. De esta manera, la semántica informacional “aumentada” con definiciones para un grupo reducido de conceptos (únicamente para los lógicos y matemáticos) puede dar cuenta de la amplia variedad de conceptos con los que pensamos.

2.2 Composicionalidad

Esta teoría resulta especialmente adecuada para dar cuenta del requisito de

composicionalidad. Éste establece que el contenido de una representación compleja debe determinarse como una función del contenido de las representaciones atómicas que la componen y las reglas de combinación sintáctica. De esta manera, la composicionalidad requiere una sintaxis y semántica combinatoria (Fodor & Pylyshyn, 1995). El atomismo informacional da cuenta de este requisito apelando a la estructura de *constituyentes* de las representaciones mentales. Fodor (2008), distingue entre las estructuras de constituyentes sintácticos y las meras estructuras de *partes*. En una estructura compleja que sólo tiene partes, cada parte realiza la misma contribución a la estructura total y no hay una forma canónica de recortar o delimitar esas partes. Las representaciones icónicas son, según Fodor, el ejemplo paradigmático de una estructura que sólo tiene partes, dado que de cualquier manera que se divida una representación icónica, se obtendrá una parte de esa representación. Por ejemplo, de cualquier modo que se recorte una foto de una tortuga, se obtiene una parte de una foto de una tortuga.

Las representaciones mentales simbólicas, en cambio, tienen una estructura en constituyentes sintácticos canónicos. Estos se diferencian de las partes en tanto no contribuyen todos de la misma manera a la representación compleja, sino que pueden tener, entre otras, relaciones de tipo predicativo. Los ejemplos paradigmáticos de este tipo de estructuras, a las que Fodor llama representaciones discursivas, son las proposiciones. Por ejemplo, en la proposición: ‘Patricio escucha música’, ‘Patricio’ y ‘escucha música’ son dos constituyentes. ‘Patricio’ es un constituyente atómico y ‘escucha música’ es un constituyente molecular, que se divide a su vez en los constituyentes atómicos ‘escucha’ y ‘música’. En cambio, ‘Patricio ... música’ es una parte de la proposición, no es un constituyente, puesto que no tiene un análisis sintáctico ni una interpretación semántica. De manera que no cualquier recorte que se haga de esta proposición delimita los constituyentes, sino que hay una estructura canónica que se delimita en relación con la contribución semántica y sintáctica que hace cada representación. Pues bien, el LDP exhibe, por definición, este tipo de estructura, puesto que sus símbolos poseen, por un lado, relaciones sintácticas, y por otro, composicionalidad semántica. Poseen, al igual que los símbolos de los lenguajes naturales, formas lógicas que se reflejan en la sintaxis, así como una relación semántica de composicionalidad entre las representaciones atómicas que dan lugar a las representaciones moleculares.

Así, el atomismo informacional puede explicar cómo podemos generar (o tener) una cantidad infinita de pensamientos a partir de un conjunto finito de representaciones, i.e. la productividad del pensamiento, por medio de la composición sintáctica y semántica de

representaciones atómicas para formar representaciones complejas. Y permite también dar cuenta de la sistematicidad, i.e. el hecho de que haya una relación entre nuestra capacidad para pensar, por ejemplo, que el ratón asustó al elefante y para pensar que el elefante asustó al ratón. No podemos tener alguno de estos pensamientos sin tener la capacidad para pensar el otro. La sistematicidad es así una relación estructural entre pensamientos y, en consecuencia, el hecho de que ambas representaciones complejas del ejemplo estén compuestas por los mismos constituyentes (junto con reglas de combinación sintáctica) permite explicar por qué no podría darse el caso de que una mente pudiera tener uno de los pensamientos, pero no ser capaz de tener el otro. De esta manera, la satisfacción del requisito de composicionalidad es otra de las grandes ventajas del atomismo informacional.

2.3 Intencionalidad

El atomismo informacional resulta especialmente atractivo por su explicación del contenido intencional de los conceptos. Tal como señalan Laurence y Margolis (1999), es una ventaja proponer que las únicas condiciones de posesión de un concepto consisten en relaciones nomológicas con el mundo y que puede haber mecanismos diferentes que permitan que se realicen estas relaciones. Al considerar que los factores epistémicos no son constitutivos de los conceptos, se puede dar cuenta de la posesión de un concepto independientemente de que se tengan creencias erróneas respecto de la propiedad que este concepto expresa o se ignoren características fundamentales de esa propiedad. Laurence y Margolis toman las enfermedades como ejemplo para ilustrar esta ventaja. En otra época se creía que la viruela era causada por espíritus malignos o “mala sangre”. Hoy decimos que en esa época estaban equivocados acerca de las causas de la enfermedad o que ignoraban características centrales tales como la forma de transmisión de la enfermedad, pero suponemos que, a pesar de las creencias erróneas, su concepto VIRUELA refería a la misma propiedad que el nuestro. La ventaja fundamental del atomismo informacional consiste en separar las creencias de la representación, al postular que la individuación del contenido de las representaciones no depende de las creencias de los individuos, sino de relaciones puramente nómicas con el mundo.

Sin embargo, como desarrollaré en la próxima sección, precisamente por excluir factores epistémicos de la determinación del contenido, la semántica informacional atomista acarrea dificultades a la hora de individuar conceptos coextensivos, i.e. conceptos que tienen el mismo contenido intencional. Éstas se deben a que, en los casos de conceptos coextensivos no se puede recurrir a relaciones nómicas mente-mundo para distinguirlos, puesto que la

propiedad con la que ambos están relacionados es la misma, ni tampoco a factores epistémicos puesto que estos no son constitutivos de los conceptos. Desarrollaré esta dificultad a continuación.

3. Las dificultades del atomismo

3.1 La individuación de conceptos coextensivos

Si bien la explicación del contenido intencional es una de las ventajas de esta teoría, se puede sostener que el contenido intencional no es suficiente para la individuación de conceptos, puesto que deja sin explicar los casos de conceptos coextensivos, llamados casos Frege, en los cuales se tienen dos conceptos distintos con el mismo contenido intencional, y los casos de tierra gemela, en los cuales un concepto tiene distintos contenidos intencionales en el mundo actual y en otro mundo posible.²⁹ Para explicar estos casos se apela usualmente a contenido cognitivo, adoptando una teoría del contenido de doble factor (Prinz 2002, Weiskopf 2009b).

Fodor admite que pueden darse casos de conceptos atómicos coextensivos, en los que dos conceptos tienen distintos poderes causales a pesar de referir a la misma propiedad y, en esos casos, la relación nómica con la propiedad no resulta suficiente para su individuación. Aquello que diferencia conceptos coextensivos, afirma Fodor (1999), son sus propiedades sintácticas. Dos representaciones mentales con el mismo contenido que están instanciadas de formas sintácticamente distintas son dos conceptos distintos. De esta manera, la explicación fodoriana no apela a otro tipo de contenido, como el cognitivo, para explicar estos casos, sino que sostiene que los responsables de una individuación de grano más fino que la que permite el puro contenido intencional son los vehículos. Así, para que dos conceptos sean distintos, es suficiente tanto que tengan distinto vehículo, i.e. que sean símbolos distintos del LDP, como que tengan distinto contenido intencional. De esta manera, la individuación de conceptos se da tanto a partir de la semántica (o contenido) como de la sintaxis (o forma).

Este modo de individuación enfrenta determinados problemas dada la concepción particular de las propiedades de apariencia con la que la teoría fodoriana se compromete. A continuación desarrollaré estos problemas, a partir de una elucidación de en qué consiste que

²⁹ No me ocuparé de los casos de tierra gemela, aunque vale la pena señalar que el tratamiento de Fodor (1994) de estos casos resulta adecuado. Fodor señala, en primer lugar, que ciertos casos de tierra gemela son nomológicamente imposibles. Por ejemplo, por los conocimientos de química disponibles, la única propiedad que tiene las características por medio de las cuales nos enlazamos con AGUA es H_2O y no XYZ (la sustancia de la que estaría compuesta el agua en la tierra gemela). Aunque esto no descarta todos los casos de tierra gemela sino sólo los del estilo H_2O/XYZ , Fodor propone tratar al resto como accidentes y considero que ésta es una estrategia adecuada.

dos representaciones mentales con el mismo contenido puedan instanciarse de formas sintácticamente distintas. Para esto abordaré dos cuestiones, cuya relación, a pesar de no ser explícita en los textos fodorianos, exhibe las dificultades que la presente teoría tiene para la individuación de conceptos coextensivos. En particular, me ocuparé de los Modos de Presentación (MDP) y de la determinación del contenido de los conceptos que covarían con propiedades de apariencia.

El tratamiento tradicional de los conceptos coextensivos apela, para su distinción, a sus MDP, i.e. objetos abstractos que nuestra mente capta y corresponden a la manera en que los objetos se presentan al pensamiento (Frege, 1969) o, en términos fodorianos, a cómo las cosas impactan en nuestra mente. Contra la estrategia fregeana, Fodor (1999) sostiene que los MDP no pueden ser objetos abstractos, sino que deben estar “en la cabeza”, i.e. deben ser particulares mentales físicamente instanciados. Para esto ofrece dos argumentos. Primero, si los MDP fueran objetos abstractos que la mente capta, según Fodor debería ofrecerse un argumento independiente para mostrar que existe una única manera de captar cada MDP, lo cual en la teoría fregeana sólo se establece por estipulación. Si hubiera una proliferación de maneras de captar cada MDP, esto resultaría o bien en una proliferación de conceptos (uno por cada manera de captar cada MDP) o bien en una asociación de un concepto con varias maneras distintas de captar un MDP. Si sucediera esto último, los MDP fallarían en la individuación de conceptos.

Segundo, Fodor sostiene que lo único que puede jugar un rol causal en los procesos mentales son las entidades físicamente instanciadas. En consecuencia, si se tiene como objetivo que los MDP puedan ser causas o efectos de procesos mentales y así funcionar como aquello que permite distinguir (en función de sus consecuencias causales) entre distintos conceptos coextensivos y entre, e.g., las actitudes proposicionales de Edipo como Deseo casarme con Yocasta y Deseo casarme con mamá, entonces los MDP deben estar físicamente instanciados. Así, Fodor (1987, 1994) propone analizar las actitudes proposicionales como relaciones entre tres elementos: un individuo, una proposición, i.e. un determinado contenido, y un MDP, i.e., en términos fodorianos, un vehículo que es una oración del LDP. La propuesta de Fodor consiste, de este modo, en “materializar” los MDP, de modo que estos pasan a estar físicamente instanciados en la cabeza como vehículos sintácticos de los conceptos. De esta manera, lo que distingue a los conceptos coextensivos es su MDP o vehículo (Fodor 1994, 1999). Decir que dos conceptos distintos son coextensivos significa que el mismo contenido intencional (una propiedad en el mundo) es transportado por dos MDP o vehículos sintácticos distintos, de modo que la misma propiedad está asociada

nómicamente con distintos símbolos del LDP.

Ahora bien, si los MDP son las maneras en que concebimos las cosas y estos están en la cabeza en tanto vehículos sintácticos de los conceptos, entonces, por ejemplo, la sintaxis de YOCASTA equivale a la manera en que la mente es impactada por la propiedad instanciada en el mundo que causa el contenido de YOCASTA. Y la sintaxis de MAMÁ equivale a la manera de la mente de ser impactada por aquello en el mundo que causa el contenido de MAMÁ.³⁰ Pero, la propiedad instanciada en el mundo que es causa de YOCASTA y MAMÁ es la misma.³¹ En consecuencia, resta explicar cómo a partir de experiencias con una misma propiedad pueden adquirirse conceptos distintos y por qué distintos vehículos covarían con instancias de la misma propiedad. La clave para esta explicación podría estar en los mecanismos de adquisición de conceptos. Fodor propone que hay diversos mecanismos de acceso semántico, entre los cuales se cuenta la percepción, que sirven como mediadores en la adquisición de conceptos, pero que no son constitutivos de estos últimos. En consecuencia, el hecho de que intervenga uno u otro mecanismo en la adquisición, no debería resultar en la adquisición de conceptos distintos. En el marco de la teoría fodoriana no puede apelarse a los mecanismos de adquisición para explicar el hecho de que una misma propiedad cause dos conceptos atómicos distintos. Considero, como intentaré mostrar a continuación, que esto conduce a un dilema que la teoría de Fodor no puede resolver sin renunciar a algún aspecto de su metafísica de los conceptos.

Si dos conceptos son coextensivos, esto significa que portan el mismo contenido. En el caso de Edipo, YOCASTA y MAMÁ portan la misma información acerca de un individuo. Si se los diferencia por su sintaxis, hace falta explicar cómo es que la información que portan se vincula con dos vehículos distintos. El caso de los conceptos coextensivos requiere o bien que una propiedad se relacione a través de una única ley psicofísica con dos conceptos distintos, o bien que haya más de una ley en relación a la misma propiedad. En el primer caso, se trataría de leyes como: ‘Los encuentros con instancias de *Yocasta* disparan los conceptos YOCASTA y MAMÁ’ o ‘La propiedad *ser Yocasta* resuena en nuestras mentes

³⁰ Me referiré siempre a MAMÁ, en lugar de MI MAMÁ o LA MADRE DE EDIPO, dado que pretendo enfocar el problema desde el punto de vista de Edipo, para el cual MAMÁ es sin dudas un concepto léxico atómico que funciona como un nombre. Tomar MI MAMÁ o LA MADRE DE EDIPO daría lugar a lecturas distintas, si se los considerara conceptos complejos en lugar de atómicos. Por otra parte, Fodor (2008, p. 65) señala que si bien podría entenderse a los nombres como descripciones abreviadas (en cuyo caso no habría problemas para la individuación), por los menos algunos nombres propios no son descripciones, sino conceptos atómicos. De esta manera, el ejemplo de YOCASTA y MAMÁ, puede tratarse como un caso de conceptos atómicos coextensivos.

³¹ En este ejemplo, la propiedad instanciada que causa ambos conceptos es la propiedad de *ser determinado individuo*. Podría hablarse directamente de *tal individuo* como el contenido con el cual se enlazan los conceptos, pero dado que Fodor habla siempre de “propiedades” mantendré la terminología. De todas maneras, considero que las consecuencias señaladas se siguen tanto para propiedades como para individuos.

YOCASTA y MAMÁ'. En el segundo caso, habría dos leyes distintas, una acerca de la instanciación de MAMÁ ante encuentros con *Yocasta* y otra acerca de la instanciación de YOCASTA ante encuentros con la misma propiedad. Ambas alternativas presentan problemas.

Si se opta por la primera opción, i.e., que una propiedad se relacione a través de una única ley con dos conceptos distintos, se estaría diciendo que tanto YOCASTA como MAMÁ caen bajo una ley que los vincula con la propiedad *ser Yocasta*.³² Esta sería una ley acerca de cómo nuestras mentes resuenan a cierta propiedad. Pero de ese modo, ante la instanciación de esta propiedad, tendría que darse la co-ocurrencia de ambos conceptos. Así, en la mente de Edipo deberían resonar ambos conceptos ante encuentros con *Yocasta*. No sería posible tener un concepto sin el otro, y esto iría en contra del atomismo conceptual, en la medida en que esta teoría implica que la posesión de un concepto atómico no depende de tener otros conceptos, sino de relaciones puramente nómicas con el mundo.

Las dificultades surgen porque no está clara la vinculación entre que una propiedad resuene de una determinada manera y que sea transportada por un vehículo determinado. Tomando YOCASTA y MAMÁ de forma independiente, la teoría permite individuar su contenido y dar cuenta de su adquisición. Una propiedad impacta la mente de Edipo y, en virtud de una ley, se dispara un concepto. De esta manera, Edipo se convierte en poseedor de una etiqueta para su enlace con dicha propiedad, por ejemplo, la etiqueta YOCASTA. La misma explicación puede formularse para MAMÁ. Pero, si el enlace se produce con dos vehículos distintos, ¿cuál de esos vehículos es el que covaría con instancias de la propiedad? Fodor debería responder que ambos. Sin embargo, la historia de Edipo sugiere más bien que depende del contexto. Pero, como reconstruí en §1, en la teoría fodoriana las condiciones de posesión e individuación de conceptos son independientes del contexto y dependen únicamente de las relaciones nomológicas mente-mundo. Entonces, si las condiciones de posesión de ambos conceptos son exactamente las mismas, el hecho de que la misma propiedad resuene en un vehículo o el otro resulta inexplicable en la teoría fodoriana.

De modo que es necesario recurrir a la segunda alternativa y postular que en estos casos hay una propiedad que se relaciona por medio de leyes distintas con conceptos distintos. Así, YOCASTA cae bajo una ley y MAMÁ bajo otra. Sin embargo, esta alternativa también presenta dificultades. Fundamentalmente, al igual que con la opción anterior, sería imposible tener un concepto sin el otro. Para evitar esta consecuencia, podría sugerirse que

³² O *ser la madre de Edipo* (ver nota 30). De todas maneras se trataría de propiedades idénticas.

hay *algo* en virtud de lo cual una propiedad se enlaza con uno u otro vehículo, i.e. que hay mecanismos distintos como mediadores de cada una de las leyes. Los mejores candidatos en la teoría para cumplir esta función son los prototipos (o más precisamente, los mecanismos inductivos de generalización que dan lugar a los prototipos). Fodor señala que hay una instancia inicial en la adquisición de conceptos en la cual (a partir de mecanismos inductivos de generalización) se forma un prototipo, que se relaciona con el concepto (o etiqueta), la que a su vez se enlazaría con la propiedad correspondiente. De este modo, la adquisición o formación de prototipos es un mecanismo más de acceso semántico que funciona como mediador en la adquisición de conceptos (Fodor 2001, 2008). Podría ocurrir que distintas experiencias con la misma propiedad dieran lugar a prototipos distintos y, de esta manera, a conceptos distintos. Pero si el hecho de adquirir distintos prototipos fuera lo que causa la adquisición de conceptos distintos, los conceptos estarían en una relación de covariación con los prototipos, y no con las propiedades. Los prototipos serían así los que determinan la instanciación de uno u otro concepto. De este modo, se estaría estableciendo una relación más fuerte entre los prototipos y los conceptos que la de ser los primeros meros disparadores de los segundos. Ahora bien, esta explicación resulta difícil de compatibilizar con la teoría fodoriana, dado que supone una relación constitutiva entre un prototipo y un concepto, lo cual Fodor (1999, 2008) rechaza.

De esta manera, la individuación de conceptos coextensivos presenta un problema que es a la vez semántico y sintáctico. Primero, y como sugiere el dilema presentado, si se postula una diferencia puramente sintáctica, ésta resulta difícil de explicar a partir de relaciones causales nomológicas como únicas asignadoras de contenido. Segundo, las dificultades para explicar esa diferencia sintáctica incluso ponen en duda si dada esta teoría pueden darse conceptos atómicos coextensivos. Parece una contradicción que una propiedad de apariencia pueda impactar en la mente en más de un modo. A partir de la definición de las propiedades de apariencia, podría parecer hasta necesario que *ser Yocasta* y *ser la madre de Edipo* fueran propiedades de apariencia distintas. Si ser una propiedad de apariencia consiste en resonar en la mente de una determinada manera, entonces propiedades que generen distintos impactos deberían ser distintas propiedades de apariencia. Así, el disparo de conceptos sintácticamente distintos quedaría explicado por tratarse de efectos de propiedades distintas.

Sin embargo, si esta fuera la explicación, i.e. si el disparo de vehículos distintos fuera una consecuencia de distintas maneras de resonar en la mente, entonces YOCASTA y MAMÁ ya no serían coextensivos. Ambos conceptos portarían información acerca de propiedades de apariencia distintas, con lo cual la diferencia entre ellos no sólo sería

sintáctica sino también semántica.³³ El tratamiento de Fodor (1994, 2008) del caso de Edipo y de los nombres propios coextensivos en general, muestra que esta multiplicación de las propiedades de apariencia no es lo que tiene en mente. Fodor afirma que en estos casos los conceptos tienen el mismo contenido y la diferencia entre ellos es puramente sintáctica. De modo que, si se los entiende como conceptos del mismo tipo que PICAPORTE, la propiedad de apariencia a la que refieren tiene que ser la misma.

El problema al intentar explicar estos casos es el de la relación entre la semántica y la sintaxis de los conceptos. Una propiedad de apariencia consiste en resonar en la mente de una determinada manera. Esto parece ser una cuestión semántica/epistémica. El impacto de las propiedades en nuestras mentes parece ser algo del orden de cómo concebimos las cosas. Ahora bien, puesto que Fodor no considera los factores epistémicos como constitutivos de los conceptos, la manera en que las propiedades resuenan en la mente debe reflejarse en la sintaxis. Sin embargo, no queda claro que la sintaxis pueda recoger las diferencias epistémicas que permiten individuar este tipo de conceptos sin apelar a los mecanismos de acceso semántico como los que determinan cómo nos impactan las propiedades y con qué propiedad covaría cada concepto.

Fodor da buenos argumentos por los cuales los factores epistémicos no deben ser parte constitutiva de los conceptos. Desplazar este tipo de factores permite, entre otras cosas, preservar la publicidad y asegurar la composicionalidad semántica de los conceptos, así como fijar la referencia de los conceptos independientemente de creencias que pueden resultar falsas. Como consecuencia, la semántica informacional atomista, junto con un mecanismo de adquisición causal-bruto sustentado por conexiones nomológicas, debería hacer todo el trabajo en las explicaciones tanto de la metafísica de los conceptos como de su adquisición. Sin embargo, los casos Frege aparecen como objeciones a esta teoría y parecen introducir nuevamente los factores epistémicos que se intentaba desplazar, puesto que gran parte del peso explicativo recae sobre los mecanismos de acceso semántico. Más aun, Fodor (2008, p. 75) sugiere que dos conceptos atómicos coextensivos podrían tener historias ontogenéticas distintas. Esto torna a la teoría más vulnerable a la crítica presentada, puesto que, si dos historias ontogenéticas distintas que involucran a la misma propiedad en el mundo dan lugar

³³ Estas dificultades podrían deberse a que no sea correcto interpretar a *ser determinado individuo* como una propiedad de apariencia. En *El Olmo y el Experto*, Fodor aclara que no sostendrá una semántica informacional “full-blown” para los nombres propios (Fodor 1994, p. 106). Sin embargo, en *Conceptos*, la hace extensiva a la mayoría de los conceptos léxicos: señala que la metafísica de PICAPORTE es la misma para “todos los conceptos que no son lógico-matemáticos y no expresan clases naturales” (Fodor 1999, p. 195). Como consecuencia, puede tratarse a los conceptos YOCASTA y MAMÁ como conceptos que refieren a propiedades de apariencia.

a distintos conceptos, esto apunta hacia los mecanismos de acceso semántico como los que explican, en última instancia, qué conceptos tenemos.

Un problema relacionado surge al intentar explicar la adquisición de conceptos adhiriendo a la tesis de que los mecanismos de acceso semántico son únicamente mediadores, como desarrollaré en la próxima sección.

3.2 Adquisición

Según la teoría fodoriana, los vehículos simbólicos son innatos, pero no sus contenidos. Esto quiere decir que hay vehículos representacionales simbólicos que están disponibles de manera innata para adquirir su contenido intencional a partir de encuentros con propiedades. Como vimos en §1, Fodor sostiene que los conceptos atómicos adquieren su contenido en virtud de la existencia de leyes psicofísicas que conectan instancias de propiedades en el mundo con instancias del concepto correspondiente (Fodor 1999). Una explicación usual de la adquisición de conceptos sostiene que se trata de un aprendizaje y que consiste en la formulación y testeo de hipótesis. Según esta postura, la relación entre la experiencia y los conceptos es evidencial: la experiencia sirve como confirmatoria o disconfirmatoria de hipótesis como ‘Eso es un pino’.³⁴

Fodor (1999, 2008) argumenta que si se entiende el aprendizaje de ese modo, surgen fundamentalmente dos inconvenientes: (i) los conceptos no son hipótesis, sino los constituyentes de las hipótesis, de modo que se necesita de todas maneras una base de conceptos no aprendidos (innatos) para poder formular las hipótesis en cuestión y (ii) dado que los conceptos no son hipótesis, lo que se adquiere al confirmar hipótesis son creencias constituidas por conceptos, no conceptos en sí. Por estos motivos, Fodor niega que la adquisición de conceptos sea un aprendizaje y sostiene, en cambio, que es un proceso causal bruto, i.e. un mecanismo de disparo del concepto ante el estímulo correspondiente en el mundo.³⁵

Ahora bien, hay de hecho una relación innegable entre la experiencia y los conceptos

³⁴ Los enfoques de prototipos, por ejemplo, adoptarían mecanismos de adquisición de este tipo, dado que los prototipos se forman por procesos de inferencias estadísticas a partir de la detección de ocurrencia conjunta de propiedades. Así, los prototipos se forman por inducción a partir de la experiencia y, en este sentido, hay una relación de evidencia entre la experiencia y la estructura interna de los prototipos.

³⁵ Muchas de las teorías que postulan mecanismos de aprendizaje que involucran la formulación y testeo de hipótesis ofrecen respuestas a estas dos críticas. Con respecto a la primera, en pocas palabras, la aceptan, y postulan una base de conceptos innata. Y, con respecto a la segunda, puesto que este tipo de posturas atribuyen una estructura interna a los conceptos, las creencias adquiridas por confirmación de hipótesis forman parte de tal estructura conceptual (e.g. Carey 2009). De este modo, para estos enfoques, adquirir creencias en algunos casos equivale a adquirir conceptos.

(aunque no sea una relación de evidencia) y esto es algo que requiere explicación. Hay una relación entre las experiencias con picaportes, por ejemplo, y la adquisición de PICAPORTE: adquirimos el concepto PICAPORTE a partir de experiencias con picaportes y no a partir de experiencias con frutillas, guitarras o licuadoras. Este es el problema que Fodor (1999) denomina picaporte/PICAPORTE. Pues bien, según Fodor, los conceptos no son aprendidos a partir de la experiencia, sino que son causados por ella. Para explicar esta relación, tal como desarrollé en §1, Fodor propone que la propiedad *ser picaporte* es una propiedad dependiente de la mente, en tanto es una propiedad que resuena en mentes como las nuestras causando que se instancie el concepto PICAPORTE. Sin embargo, para que el mecanismo causal bruto funcione, es necesario que nuestras mentes puedan entrar en contacto con las propiedades en el mundo. Para esto, como anticipé más arriba, disponemos de ciertos mecanismos de acceso semántico, entre ellos la percepción. Otros mecanismos de este tipo pueden ser, por ejemplo, las inferencias teóricas o los instrumentos tecnológicos como telescopios y radares. Estos mecanismos se introducen en la teoría como mediadores en el vínculo causal entre las propiedades y los conceptos.

Los mecanismos de acceso semántico permiten que la mente entre en relación con las propiedades en el mundo, y así, posibilitan el enlace con las mismas. Como vimos, la relación entre los conceptos y las propiedades es de tipo nomológico. Sin embargo, esto no implica que esa relación deba ser básica, i.e. inmediata. Ese sería el caso si la mente humana pudiera entrar directamente en contacto con las propiedades, sin ningún tipo de mediación. Pero es un hecho acerca de los seres humanos que para que nuestras mentes entren en contacto con el mundo necesitamos algún tipo de mecanismo de acceso. Esta es la función que cumplen los mecanismos de acceso semántico.³⁶ Así, las relaciones nómicas que confieren contenido a los conceptos se sustentan en estos mecanismos, pero ellos no son constitutivos del contenido de las representaciones mentales. Dada esta concepción, no es necesario tener ningún mecanismo de acceso semántico en particular, sino tener *algún* mecanismo, y las diferencias de acceso entre distintos individuos a la propiedad en el mundo no atentan contra la publicidad de los conceptos, i.e. el hecho de que todos accedamos a sus contenidos. Esto es así porque, según el atomismo informacional, los mecanismos de acceso semántico no son constitutivos del contenido, de modo que distintos mecanismos no modifican la información

³⁶ Esta función de mediación no es metafísicamente necesaria sino contingente. De manera que podríamos haber tenido un acceso semántico inmediato a las propiedades, o haber tenido otros tipos de mecanismos. Asimismo, es un hecho contingente acerca de nuestra constitución como seres humanos que la percepción sea el medio central para el acceso semántico (Fodor 1999).

que transportan los conceptos.³⁷ Lo único que determina el contenido es una relación nomológica con el mundo, la cual, aunque esté mediada por distintos mecanismos, se supone que es independiente de ellos.

Ahora bien, entre los mecanismos mediadores se incluye también la formación (o el aprendizaje) de prototipos. Fodor (2001, p. 144) explicita el rol de los prototipos al explicar las propiedades dependientes de la mente: “*ser picaporte* es tener la propiedad con la cual mentes como las nuestras se enlazan como consecuencia del tipo de experiencias a partir de las cuales nuestro tipo de mente aprende el prototipo de picaporte”. Según Fodor, la adquisición de PICAPORTE se produce en dos fases: por un lado, a partir de experiencias con picaportes se forma un prototipo (por mecanismos inductivos de generalización estadística). Y, por otro lado, opera una función de asociación de prototipos a conceptos, por la cual se produce el enlace entre la propiedad *picaporte* en el mundo y el vehículo PICAPORTE. De esta manera los prototipos funcionan como mediadores en la adquisición de conceptos, i.e., funcionan como mecanismos de acceso semántico, pero, al igual que el resto de estos mecanismos, no son constitutivos de los conceptos (Fodor 2001, 2008). Así se evitan los problemas que, como hemos visto en el capítulo 2, presentan los prototipos, en particular, los relativos a la composicionalidad.

Sin embargo, esta postulación de dos fases de adquisición no explica cómo se produce el pasaje de prototipos a conceptos, específicamente, cómo es que se pasa del *aprendizaje* del prototipo a la *adquisición* del concepto. Fodor (2008) considera que la adquisición de un concepto, i.e. el pasaje del prototipo al concepto, no puede ser un proceso inferencial, y muestra que ninguna alternativa que pudiera instanciar dicho proceso resulta viable. Una posibilidad sería que se compare un prototipo con conceptos candidatos, pero de esta manera se presupone que ya se tienen los conceptos candidatos cuando lo que se quiere explicar es, precisamente, la adquisición de conceptos. Esta alternativa es evidentemente circular. Una segunda opción sería que exista un patrón de inferencia de acuerdo al cual se pueda inferir un concepto dado su prototipo. Respecto de esta posibilidad, Fodor objeta que los conceptos y sus prototipos correspondientes pueden relacionarse de muchas maneras diferentes. Una relación puede ser la similitud, por ejemplo, para el caso de las instancias que caen bajo PARAGUAS, pero esta relación no se aplica en el caso del prototipo de, por ejemplo, PRESIDENTE. Un mismo prototipo podría incluso corresponder a distintos conceptos (Fodor

³⁷ Aunque podrían ser responsables de determinar qué vehículo transportará ese contenido, tal como argumenté en relación a la individuación de conceptos coextensivos.

sostiene que uno no puede simplemente observar un prototipo y saber a qué concepto corresponde).

Por estas razones, la explicación del pasaje del prototipo al concepto tiene que ser de otro tipo, no puede ser un proceso inferencial. Fodor apela a la neurología y al tipo de mente que tenemos. Nuestras mentes mapean prototipos a conceptos. Cuando aprendemos el prototipo de mesa, adquirimos MESA. Esta explicación, sin embargo, no esclarece demasiado, puesto que apela a una suerte de “armonía preestablecida” entre los conceptos y sus prototipos. Si un mismo prototipo puede corresponder a distintos conceptos, ¿qué es lo que determina el pasaje de dicho prototipo a uno u otro concepto? “Un proceso neurológico” es la única respuesta que ofrece Fodor (2008).

Fodor incluso señala que hay una disposición innata para enlazarse con determinadas propiedades al aprender determinados prototipos. De esta manera, el equipamiento innato que se requiere para adquirir conceptos incluye, por un lado, ciertos mecanismos (como los mecanismos perceptivos) que ponen a nuestras mentes en contacto con las propiedades y les permiten así resonar (Fodor 1999) y, por otro lado, una disposición innata para enlazarse con propiedades al aprender sus prototipos (Fodor 2008). Esto indica que el aprendizaje de los prototipos no es suficiente para la adquisición de conceptos, sino que se requiere además una disposición innata para conectar esos prototipos con los conceptos correspondientes. Ahora bien, según Fodor, esta disposición no es intencional, en el sentido de que no involucra una relación evidencial racional entre los prototipos y los conceptos, sino que se trata de una disposición neurológica. Por un proceso neurológico los prototipos median la vinculación con las propiedades a las que nuestros conceptos refieren. Sin embargo, si un mismo prototipo puede corresponder a distintos conceptos, y, como afirma Fodor, “mirar” un prototipo no resulta suficiente para saber a qué concepto corresponde, no queda claro cómo el proceso neurológico al que Fodor alude, vincula los conceptos con sus correspondientes contenidos, sin que “lo confundan” los prototipos. Si no se detallan las características del mecanismo que vincula prototipos y conceptos, esta versión no parece *explicar* cómo se produce la adquisición (Landau 2000).

3.3 Categorización

La otra gran desventaja del atomismo informacional es su dificultad para explicar la categorización, debido a dos motivos. Primero, una explicación adecuada de la categorización debe acomodar los distintos efectos asociados a este proceso reportados en la literatura, presentados en los capítulos 2 y 3, para lo cual las mejores explicaciones disponibles apelan a

algún tipo de estructura conceptual interna. Si bien ninguno de los enfoques presentados en esos capítulos consigue acomodar todos los efectos reportados, cada uno resulta adecuado para dar cuenta de algunos de ellos. Cada efecto parece revelar alguna característica de la estructura conceptual. El atomismo tiene como única respuesta disponible sostener que estos efectos se explican por conocimiento asociado a cada categoría, que no es constitutivo de los conceptos. De hecho, Fodor no considera que una teoría metafísica de conceptos deba explicar la categorización, sino que sostiene que esta tarea corresponde a la psicología cognitiva (Fodor 2000b). Sin embargo, si las mejores teorías psicológicas disponibles postulan representaciones mentales internamente estructuradas para explicar estos procesos que resultan en la asignación de un concepto a una instancia, resultaría interesante ofrecer una explicación del vínculo entre esas estructuras y las etiquetas que postula el atomismo informacional. En cualquier caso, la postulación de representaciones mentales con una estructura interna, ya sea prototípica, de ejemplares o teórica, implica ventajas explicativas en relación a los procesos de categorización, con las que no cuenta el atomismo. Retomaré esta crítica más adelante.

Segundo, el atomismo forma parte de una teoría general de la mente, que tiene dificultades para dar cuenta de procesos en los que intervienen creencias y conocimiento de trasfondo, como probablemente sea el caso para la categorización. Fodor sostiene que *pensar es computar*. Esto significa que los procesos de pensamiento consisten en transiciones causales entre símbolos: pensar es pasar de una representación mental a otra de manera ordenada, siguiendo determinados algoritmos, en virtud de las propiedades sintácticas o formales de esas representaciones mentales. Por ejemplo, sobre una representación compleja como HACE FRÍO Y ESTOY LEJOS DE CASA puede operar una transformación que dé lugar a HACE FRÍO, porque esa representación tiene la forma P&Q y sobre representaciones de esta forma puede operarse una transformación que da lugar a Q. Los procesos mentales computacionales operan así sobre la estructura sintáctica de las representaciones mentales y son sensibles *únicamente* a esta estructura. Así, las transformaciones de este tipo dependen sólo de las propiedades sintácticas de las representaciones, pero tienen la virtud de preservar de manera confiable los valores semánticos que portan los vehículos representacionales sintácticos. Esto es así porque, como vimos en §1, la sintaxis sirve de vehículo a los contenidos, i.e. codifica los contenidos intencionales y “refleja” así relaciones semánticas.

Fodor sostiene también que las explicaciones psicológicas apelan a los contenidos de los estados mentales para explicar la conducta. Así, para explicar el comportamiento, la psicología formula generalizaciones legaliformes del tipo: si una persona *desea* P y *crea* que

P sólo si Q, entonces, *ceteris paribus*, esa persona hará Q (o intentará llevar a cabo Q) (Fodor, 1994). Por ejemplo, si una persona cree que trae mala suerte pasar por debajo de una escalera y desea evitar la mala suerte, entonces, al caminar por la calle, evitará a toda costa pasar por debajo de una escalera. Este tipo de explicación se basa en el contenido del deseo y la creencia para predecir una cierta conducta. Sin embargo, el contenido por sí mismo no tiene poder causal. La manera en que los contenidos pueden causar comportamientos u otros estados mentales es por el hecho de estar “reflejados” en la sintaxis. Así, los procesos mentales operan en virtud de las propiedades sintácticas, pero éstas, a su vez, transportan el contenido, de modo que sirven de vínculo entre el contenido y las relaciones causales que permiten explicar la conducta. De esta manera, las transformaciones sintácticas “respetan” el contenido (Fodor 1987). Además, “en tanto que la sintaxis codifica aspectos semánticos que determinan los roles inferenciales de los símbolos, los procesos computacionales respetan la coherencia semántica del pensamiento. Así, son posibles mecánicamente razonamientos semánticamente coherentes (y que preservan la verdad)” (Skidelsky en prensa, p. 85).

Ahora bien, las propiedades sintácticas que gobiernan los procesos computacionales son propiedades locales de las representaciones mentales. Esto significa que son propiedades que dependen sólo de las características intrínsecas o esenciales de las representaciones mentales en cuestión, específicamente, dependen de las partes que constituyen una representación mental y de cómo esas partes están relacionadas entre sí. Como señala Fodor (2000a, p. 20): “Para ver cuál es la estructura sintáctica de una oración no es necesario mirar “fuera” de ella...”. En este sentido, la sintaxis es una propiedad de las representaciones mentales que no depende de ninguna relación con otras representaciones por fuera de la que está en cuestión (así como no depende de que se tengan otras representaciones mentales en absoluto). De modo que las propiedades locales son esenciales para las representaciones mentales, si cambiara alguna de ellas estaríamos en presencia de una representación diferente. Así, la identidad de las representaciones está definida sobre propiedades locales. Las propiedades globales, en cambio, se relacionan con elementos por fuera de la estructura constituyente de una representación mental, y así dependen de sistemas completos de representaciones o creencias, de allí su denominación. Por ejemplo, la relevancia es una propiedad global. No se puede determinar la relevancia de una representación atendiendo únicamente a sus características intrínsecas o estructura constituyente, sino que esta propiedad se determina de alguna manera por su relación con el resto de las representaciones mentales. Si se desea evaluar la relevancia de una determinada representación para la confirmación o no de una hipótesis, ésta sólo se podrá determinar en relación con un conjunto

de creencias de trasfondo. Otros ejemplos de propiedades globales son la simplicidad y la plausibilidad, las cuales también se determinan en función de conjuntos de creencias (potencialmente conjuntos completos de creencias) y serían imposibles de determinar atendiendo sólo a las propiedades intrínsecas de las representaciones mentales.

Pues bien, los procesos computacionales pueden operar únicamente sobre las propiedades locales, dado que están gobernados exclusivamente por la sintaxis de las representaciones (son sensibles únicamente a la identidad de las representaciones y su estructura constituyente). Por este motivo, una TRCM sólo podrá explicar los procesos que dependen de este tipo de propiedades (Fodor 2000a, 2008). Así, la TRCM resulta ideal para explicar los procesos de razonamiento deductivo, puesto que estos dependen únicamente de la forma lógica de los pensamientos, la cual está plasmada en la estructura sintáctica de las representaciones. Por ejemplo, deducir que Charly va en avión, a partir de las creencias de que Charly va en tren o en avión y que no va en tren, no requiere considerar ninguna propiedad por fuera de las propiedades sintácticas de las representaciones que sirven de vehículo a cada una de las creencias.

Ahora bien, no todos los procesos mentales se pueden explicar atendiendo a propiedades locales, dado que parte de nuestra vida mental es sensible a propiedades globales.³⁸ Este es el caso, en particular, de los procesos de fijación de creencias empíricas, que son análogos a los de confirmación de hipótesis empíricas. Se trata de procesos inductivos o abductivos de inferencia (típicamente inferencias a la mejor explicación), que dependen de propiedades como la relevancia y la simplicidad para determinar qué se considerará como confirmatorio o no de una hipótesis. Según Fodor (1983, 2008) los procesos de confirmación de hipótesis científicas tienen dos características que los convierten

³⁸ La distinción entre procesos locales y globales (Fodor 2000a) está relacionada con (aunque no es idéntica a) la distinción entre sistemas de *input* y sistemas centrales que se plantea en la teoría modular de la arquitectura cognitiva (Fodor 1983). No es mi intención desarrollar aquí esta teoría, pero cabe aclarar brevemente las características de cada uno de estos sistemas. Los sistemas de *input* son modulares, i.e. son mecanismos computacionales especializados que están informacionalmente encapsulados y son específicos de dominio, entre otras características, y tienen la función de generar representaciones del carácter (y distribución en algunos casos) de estímulos distales. Entre ellos se incluyen el lenguaje y los sistemas perceptivos, que no son equivalentes a los sentidos, sino más específicos (incluyen, entre otros, mecanismos de percepción de color y de análisis de formas y mecanismos que detectan el ritmo de los estímulos acústicos). Los sistemas centrales, en cambio, no están informacionalmente encapsulados y no son modulares. Son los que reciben la información producida por los sistemas de *input* y tienen como función principal la fijación de creencias mediante procesos inferenciales no demostrativos. De manera que se podría pensar que mientras que los procesos modulares son completamente computacionales, i.e. operan en virtud de las propiedades sintácticas locales de las representaciones mentales, la mayor parte de los procesos centrales no parecen ser computacionales dado que operan en virtud de propiedades globales. Retomaré la distinción entre sistemas modulares y centrales en el capítulo 7, al abordar el tipo de arquitectura que parece implicar una propuesta pluralista como la que desarrollaré.

en procesos holistas. Primero, son *isotrópicos*, i.e., en principio, cualquier creencia previamente establecida puede influir en la determinación de los hechos relevantes para su confirmación. Segundo, son *quineanos*, en tanto el grado de confirmación de una hipótesis depende de propiedades globales, tales como la simplicidad y la plausibilidad. De modo que si los procesos centrales de fijación de creencia son análogos a los procesos científicos de confirmación de hipótesis, estos también serán isotrópicos y quineanos. Así, este tipo de proceso resulta ser holista, puesto que depende de propiedades globales, las cuales se basan en una relación epistémica entre *todas* o gran parte de las creencias. Este tipo de procesos no resulta explicable en el marco de una TRCM, puesto que los procesos computacionales que esta teoría postula, como vimos, dependen sólo de propiedades locales.

La moraleja es que hay procesos mentales que una TRCM no puede explicar y entre ellos se incluyen procesos fundamentales para nuestra vida mental, y nuestra vida en general, como la fijación de creencias y el aprendizaje. Este es el caso, también, de los procesos de categorización. Dado que la categorización es, según Fodor, un proceso abductivo de fijación de creencias, este proceso depende de propiedades globales. Ya sea que se postule que la categorización se lleva a cabo por un proceso de comparación de rasgos o recurriendo a información de tipo causal, como proponen, respectivamente, los defensores de las teorías de prototipos y ejemplares, y los defensores de la teoría-teoría, como se vio en los capítulos 2 y 3, será necesario delimitar qué conjuntos de rasgos o propiedades se tendrán en cuenta, i.e. cuáles serán *relevantes* para la tarea, de modo que será necesario recurrir a propiedades globales que no son parte de la estructura constituyente de las representaciones mentales. Respecto de la adquisición, si se considera que consiste en un aprendizaje que implica la formulación y testeo de hipótesis, este proceso tampoco podrá explicarse computacionalmente, puesto que sería un caso de proceso inductivo o abductivo de fijación de creencias.

Fodor (1999, p. 46) sostiene que los conceptos “son categorías y se emplean, de manera rutinaria, como tales” y que es un requisito “no negociable” de una teoría de conceptos dar cuenta de esto. Sin embargo, no se puede explicar la categorización por medio de la TRCM dado que este es un proceso central de fijación de creencias por medio de inferencias abductivas, de modo que el atomismo debería incluir otro tipo de procesos mentales que puedan operar sobre las propiedades globales de las representaciones para poder satisfacer este requisito. Por otra parte, al postular que los conceptos son atómicos, i.e. no tienen estructura interna, y que están constituidos independientemente de cualquier habilidad epistémica, no puede explicar la categorización sin recurrir a elementos no constitutivos de

los conceptos (Laurence & Margolis 1999, Prinz 2002). Desarrollaré primero esta última dificultad, y luego retomaré el problema de los procesos globales.

Para explicar la aplicación, por ejemplo, del concepto GATO, dado el atomismo informacional, se debería recurrir o bien a las mismas leyes psicofísicas que establecen las relaciones entre los conceptos y el mundo, o bien a elementos no constitutivos del concepto. Se podría defender la primera alternativa sugiriendo que, dado que las leyes psicofísicas establecen que ante instancias de gatos en el mundo se disparará el concepto GATO, el categorizar una instancia de gato como GATO, es simplemente representarlo como cayendo bajo el concepto GATO. Así, se podría explicar la categorización acudiendo a las mismas leyes psicofísicas que determinan el contenido de los conceptos. Sin embargo, podría ocurrir que estuviera oscuro, y alguien, frente a un gato, pensara LIEBRE. Podemos confundir gatos por liebres. ¿Significa esto que no poseemos el concepto GATO?

Fodor (1999) responde a esta pregunta afirmando que no existe ningún escenario tan claramente “gatuno” como para que la falla al categorizar erróneamente una instancia de gato bajo algún otro concepto implique la falta de posesión del concepto GATO. Esto es así porque Fodor diferencia entre las condiciones de posesión y las de aplicación de un concepto. Mientras que las primeras le corresponden al concepto de manera esencial, puesto que poseer un concepto no es más que poseer una instancia de una representación que está nomológicamente conectada con una instancia de una propiedad en el mundo, las segundas no. Las condiciones de posesión dependen de la metafísica de los conceptos y son esenciales en tanto dependen de las relaciones nomológicas constitutivas de los conceptos. Tener una representación nomológicamente conectada con la propiedad de *ser gato*, implica que se posee el concepto GATO. Así, las condiciones de posesión dependen de una relación semántica: la relación nomológica constitutiva del contenido de los conceptos.

En cambio, las condiciones de aplicación de un concepto dependen de relaciones epistémicas que no son constitutivas de los conceptos. Si bien la categorización establece, al igual que la representación, una relación entre conceptos y objetos, esta relación es distinta de la de representación (i.e. la relación nomológica que determina el contenido) puesto que no consiste en una mera relación causal entre un concepto y una propiedad, sino que en la categorización intervienen (en principio, de manera constitutiva) creencias, teorías y mecanismos perceptivos. Por este motivo, Fodor no puede explicar la categorización sin apelar a elementos no constitutivos de los conceptos.

Dado este escenario, reconstruiré en primer lugar la crítica al atomismo según la cual el hecho de que las características intrínsecas de los conceptos no sirvan para explicar la

categorización implica una debilidad de la teoría (Prinz 2002), pero sostendré, siguiendo a Weinberg (2003), que los *desiderata* de una teoría de conceptos imponen que se dé cuenta de estos requisitos, pero no que se lo haga apelando exclusivamente a las características intrínsecas de los conceptos. En segundo lugar, desarrollaré una sugerencia de Fodor que podría permitir articular el atomismo informacional con las explicaciones de la psicología cognitiva acerca del proceso de categorización y, por lo tanto, con explicaciones de un proceso global.

Prinz (2002) señala que no poder explicar la categorización implica una debilidad importante para el atomismo conceptual, puesto que en las teorías de la psicología cognitiva se presta especial atención a los conceptos precisamente por su rol en la categorización, i.e. por su contribución en la determinación de que un elemento pertenece a una determinada categoría. Sin embargo, Prinz admite que la no satisfacción de este *desideratum* no es suficiente para descartar la teoría, aunque la deja en una posición débil frente a otras teorías que, pudiendo explicar el resto de los *desiderata*, pueden explicar también la categorización por medio de características esenciales de los conceptos. Frente a esta crítica, Weinberg (2003) objeta que Prinz “confunde las reglas del juego”. Una buena teoría de conceptos debería satisfacer ciertos requisitos explicativos, pero esto no implica que todas las explicaciones deban basarse en las características metafísicas de los conceptos. Fodor se limita a señalar que las características metafísicas de los conceptos tienen que ser compatibles con aquello que las teorías psicológicas postulen para dar cuenta de nuestra vida mental. Así, reconoce el rol central que tiene en la vida mental la aplicación de conceptos como categorías y, por lo tanto, que aquello que postule una teoría de conceptos tiene que ser compatible con que los conceptos se apliquen cotidianamente en la categorización de objetos.

Tal como Fodor mismo sostiene: “Una teoría de conceptos tiene que explicar dos cosas: cómo los conceptos funcionan como categorías, y cómo una mente finita puede tener una capacidad conceptual infinita y sistemática” (1999, p. 149). Como sostuve en §2.2, los conceptos entendidos a la manera fodoriana permiten explicar la capacidad conceptual infinita y sistemática, dado que la atomicidad de los conceptos los vuelve especialmente adecuados para dar cuenta de la composicionalidad. Así, a partir de un conjunto finito de conceptos primitivos y reglas de combinación se pueden generar combinaciones infinitas y relacionadas entre sí. Ahora bien, respecto al requisito de explicar cómo los conceptos funcionan como categorías, la semántica informacional sólo dice que un concepto tiene como referente una clase de cosas: todas aquellas que instancian la propiedad con la que el concepto está nomológicamente relacionado. En este sentido, un concepto es una categoría.

Sin embargo, los *procesos* de categorización no pueden explicarse exhaustivamente sólo a partir de las representaciones conceptuales atómicas. Si esto fuera así, estos procesos tendrían que explicarse solamente a partir de leyes psicofísicas y esto no se sostiene. En la categorización intervienen también mecanismos perceptivos y creencias. Fodor (1999) no cree que sea posible formular una teoría de la individuación de las creencias, puesto que habría que recurrir para ello a la individuación por roles funcionales/causales, lo cual, tiene como consecuencia el holismo. Esto es así, porque para individuar por el rol funcional/causal habría que tener en cuenta la intervención de las creencias en los procesos centrales, de modo que sería necesario establecer relaciones epistémicas entre la creencia que se quiere individuar y todo el resto del sistema.

Sin una teoría semejante, tampoco se pueden especificar las condiciones de verificación de los conceptos. Fodor diferencia entre las condiciones de satisfacción y las de verificación de los conceptos. Adopta una postura atomista respecto de las condiciones de satisfacción de un concepto, pero holista respecto de las condiciones de verificación. De esta manera, los conceptos tienen esencialmente sus condiciones de satisfacción, y estas condiciones son atómicas, puesto que consisten en una relación mente-mundo que no depende de otras relaciones. Una propiedad instanciada cae bajo un concepto o no dependiendo de las relaciones nómicas. Las condiciones de verificación, en cambio, se refieren a las condiciones para *decidir* si un ítem cae bajo un concepto o no. Para cada instancia particular de aplicación de un concepto, estas relaciones están mediadas por teorías y creencias y son, por lo tanto, holistas (Fodor 1999). Estas relaciones no le corresponden al concepto de manera esencial, sino que al estar mediadas por teorías y creencias, varían de persona a persona, mientras que los conceptos son públicos y esencialmente idénticos (en tipo) en todas las mentes.

Ahora bien, a pesar de la distinción entre satisfacción y verificación, y de las dificultades para individuar creencias y, por lo tanto, para especificar condiciones de verificación, resulta explicativamente débil sostener que los conceptos se usan como categorías, pero que los procesos que permiten determinar si una instancia corresponde a una categoría, no forman parte de la teoría metafísica de los conceptos. El problema no es tanto que Fodor no puede explicar la categorización sin recurrir a mecanismos no constitutivos de los conceptos, sino que directamente excluye esta explicación de su teoría. Fodor sostiene que la aplicación de un concepto en una situación determinada, i.e. el reconocimiento de una instancia como cayendo bajo un concepto, es una forma de adquisición de una nueva creencia (por ejemplo reconocer a un gato como GATO implica formar o adquirir la creencia 'Ese es

un gato') y, por lo tanto, no es meramente el disparo de un concepto frente a una instancia.

Ante las críticas por la falta de una explicación de la categorización –por ejemplo, Landau (2000)–, Fodor (2000b) se desentiende del problema sosteniendo que no es un requisito para una teoría metafísica de conceptos explicar *cómo se aplican* los conceptos sino especificar *qué* es un concepto y cuáles son sus condiciones de posesión. Así, no es un requisito explicar qué mecanismos permiten determinar si cierto objeto cae bajo un concepto o no, sino explicar las propiedades metafísicas de los conceptos. La encargada de dar cuenta de este tipo de procesos sería la psicología cognitiva. Sin embargo, retomando las consideraciones respecto de los procesos globales o centrales, Fodor (1983) considera que los únicos sistemas que la psicología cognitiva logrará explicar satisfactoriamente serán los sistemas modulares que operan sobre propiedades locales de las representaciones mentales. De modo que la exclusión del requisito de categorización de las teorías de conceptos es consistente con su enfoque de la arquitectura cognitiva y del tipo de explicación que la ciencia cognitiva pueden formular.

Ahora bien, esta perspectiva resulta quizás excesivamente desesperanzada, dado que, como vimos en los capítulos 2 y 3, hay teorías psicológicas disponibles que intentan explicar la categorización recurriendo a representaciones de rasgos asociadas a cada categoría y, dependiendo del tipo de estructura postulada, a distintos tipos de operaciones sobre esas representaciones, como procesos de comparación de rasgos o de inferencias a la mejor explicación. Pues bien, Fodor (1999) sostiene que la metafísica de los conceptos debe ser compatible con las explicaciones de la psicología acerca de las funciones de los conceptos, pero para que el atomismo informacional sea compatible con éstas, sería necesario explicar la articulación entre las representaciones conceptuales atómicas y la información asociada a ellas (los rasgos que determinan que esa categoría se asigne a un elemento o no). Fodor (1999) acepta que los conceptos tienen prototipos y que el aprendizaje de los prototipos es una etapa de la adquisición de conceptos. De modo que se podría recurrir a estos prototipos asociados a los conceptos para explicar la categorización. Esta sugerencia no parece descabellada puesto que Fodor mismo admite que los prototipos resultan adecuados (o *bastante* adecuados) para explicar la categorización (Fodor & Lepore 1996).

Por otra parte, Fodor (2008) también recurre a la metáfora de los archivos mentales para sugerir un modo en que los conceptos pueden estar asociados a conocimiento acerca de las propiedades que representan. Sostiene que parte de nuestra mente podría estar organizada en forma de archivos mentales etiquetados por conceptos y se refiere a estos archivos como conjuntos de creencias en LDP. La información que sirve para explicar la categorización

podría ser parte de estos archivos y así estar codificada o transportada por oraciones en mentalés. Esto parece plausible, puesto que, al igual que los teóricos de las teorías de prototipos que ofrecen descripciones lingüísticas para caracterizar a este tipo de estructuras, se podría postular una descripción análoga en mentalés. Podría interpretarse la metáfora de los archivos mentales como un elemento que permitiría articular la teoría fodoriana con las explicaciones psicológicas de la categorización. En el capítulo 7 desarrollaré una propuesta de este tipo. La articulación del atomismo informacional con determinados modelos de la psicología cognitiva que explican la categorización, caerá inevitablemente por fuera de la TRCM. Sin embargo, considero que esto no es una desventaja, sino que implicaría un avance en la dirección de explicar los procesos centrales.

El análisis llevado a cabo en estos primeros capítulos evidencia la necesidad de incorporar distintos tipos de estructuras conceptuales, si se pretende dar cuenta de los requisitos explicativos fundamentales. En el próximo capítulo, reconstruiré y evaluaré el neo-empirismo de conceptos, en particular la teoría de proxitipos de Prinz que, a partir de una conclusión similar, propone que los conceptos pueden tener una estructura interna prototípica, de conjuntos de ejemplares o de tipo teórico. Esta teoría puede ser considerada híbrida, en tanto sostiene que los conceptos tienen partes con distintos tipos de estructuras. Al incorporar los principales tipos de estructuras semánticas, esta teoría promete heredar las ventajas explicativas de cada una de ellas. Sin embargo, una de las tesis principales de la teoría concierne a los vehículos conceptuales y postula que estos son representaciones perceptivas (o copias de ellas). Así, en relación a los vehículos conceptuales, la teoría de proxitipos conserva el principio de uniformidad. Como sostendré en el próximo capítulo, la tesis de que todos los vehículos conceptuales son representaciones perceptivas acarrea múltiples dificultades y limita el poder explicativo de la teoría.

Capítulo 5

El neo-empirismo de conceptos

En los capítulos 2 y 3 me ocupé de distintos enfoques relativos a la estructura semántica de los conceptos. En el capítulo 4 abordé el debate en torno a los vehículos de los conceptos, al reconstruir y evaluar las tesis principales del atomismo conceptual fodoriano, según el cual los vehículos conceptuales son símbolos del LDP. En este capítulo me ocuparé de un enfoque alternativo respecto de los vehículos, el neo-empirismo de conceptos, según el cual los conceptos son copias o combinaciones de copias de representaciones perceptivas. Abordaré la teoría de proxitipos de Prinz, una de las principales representantes de este enfoque. En cuanto a los vehículos, esta teoría adhiere a la tesis de la uniformidad, en tanto afirma que todos los vehículos conceptuales son de tipo perceptivo. En cuanto a la estructura semántica conceptual, en cambio, esta teoría adopta una versión híbrida o pluralista, puesto que sostiene que los conceptos pueden tener distintos tipos de estructuras, en particular, pueden ser prototipos, conjuntos de ejemplares, o teorías. En §1 presentaré las tesis relativas a los vehículos y a la estructura semántica. En §2 sostendré que estas dos tesis le dan a la teoría de proxitipos ciertas ventajas. La apelación a representaciones perceptivas permite establecer una conexión entre la percepción y la cognición que resulta atractiva para la explicación de la adquisición de los conceptos. Además, la tesis de que los procesos conceptuales involucran a los mismos mecanismos que los procesos perceptivos cuenta con apoyo empírico y hay una serie de efectos observados en tareas de categorización para los cuales la mejor explicación implica postular vehículos de tipo perceptivo. Por otro lado, la tesis pluralista relativa a la estructura conceptual, le otorga la ventaja de poder dar cuenta de los efectos asociados a cada una de las estructuras que incorpora. Sin embargo, en §3, mostraré que la tesis de que todos los conceptos son copias o combinaciones de copias de representaciones perceptivas, acarrea dificultades. Prinz defiende el neo-empirismo de conceptos a partir de un argumento de parsimonia, según el cual las representaciones perceptivas son suficientes para dar cuenta de todos los fenómenos que una teoría de conceptos debería explicar, de modo que postular representaciones amodales (como las del atomismo conceptual). En este capítulo mostraré que el argumento de parsimonia no logra lo que se propone, puesto que apelando únicamente a representaciones perceptivas el neo-empirismo no consigue dar cuenta de los fenómenos de contenido ni alcance. En este sentido, sostendré que tanto para la individuación del contenido de los conceptos, como para la representación de conceptos abstractos y teóricos es necesario recurrir a representaciones

amodales. Por otra parte, señalaré que la postulación de prototipos, ejemplares y teorías como posibles estructuras conceptuales acarrea ciertas dificultades con respecto al requisito de composicionalidad.

1. El neo-empirismo de conceptos

Una de las tesis tradicionales del empirismo, afirma que todo lo que está en el pensamiento proviene de los sentidos. Esta tesis se ha defendido de distintas maneras, con interpretaciones en algunos casos radicales. En general, los enfoques empiristas coinciden en otorgar prioridad a la experiencia sensible, ya sea como fuente de conocimiento y justificación de creencias, o como fuente de las representaciones conceptuales. Hacia fines de la década de 1990 cobró fuerza en los ámbitos de la psicología cognitiva y la neuropsicología una interpretación de esta tesis como una tesis en torno a los vehículos del pensamiento, según la cual los conceptos están vehiculados en el mismo formato que las representaciones que se activan en la percepción. Según esta vertiente, a la que se denomina “neo-empirismo” de conceptos, los sistemas de representación conceptual comparten mecanismos de procesamiento con los sistemas perceptivos (Barsalou 1999, Barsalou *et al.* 2003, Barsalou, Pecher, Zeelenberg, Simmons & Hamann 2005, Prinz 2002).

En lo que sigue presentaré las tesis principales del neo-empirismo de conceptos, i.e. la teoría de Prinz, quien se basa en los trabajos de Barsalou y sus colegas. Los principales representantes contemporáneos del enfoque. El neo-empirismo de conceptos defiende fundamentalmente una tesis acerca de los vehículos conceptuales: afirma que los conceptos están vehiculados por representaciones perceptivas o modales, i.e. que el vehículo que transporta el contenido de los conceptos es el propio de cada uno de los sistemas perceptivos. Prinz caracteriza a estos últimos como sistemas de *input* dedicados. Se trata de sistemas que presentan un conjunto distintivo de representaciones y operaciones alojados en una vía neural propia y que responden a *inputs* específicos de dominio (por ejemplo, la visión responde a longitudes de onda, mientras que el olfato responde a las formas de las moléculas). Las representaciones perceptivas son los *outputs* de cada uno de estos sistemas, son “representaciones que pueden ser causadas por la estimulación del entorno junto con mecanismos en los sistemas de *input*” (Prinz 2002, p. 197).

Esta caracterización de las representaciones conceptuales se opone a aquella que las concibe como amodales o simbólicas, tal como el LDP (Fodor 1999). Como vimos en el capítulo anterior, las representaciones amodales son análogas a los símbolos de un lenguaje natural. Por un lado, al igual que los lenguajes naturales, presentan una semántica y una

sintaxis combinatoria (Fodor 1975). Por otro lado, estas representaciones tienen una relación arbitraria con las propiedades que representan, en tanto no exhiben ningún tipo de isomorfismo con aquello que representan y su “estructura interna no guarda correspondencia con los estados perceptivos que las produjeron” (Barsalou, 1999, p. 578). Las similitudes entre las propiedades formales de dos representaciones amodales no necesariamente implican que haya similitudes correspondientes entre aquello que representan. Por ejemplo, en un sistema amodal, el vehículo de AZUL no necesariamente resultará más similar al de VERDE que al de ROJO (Barsalou, Solomon & Wu 1999). En este sentido, las representaciones amodales funcionan como etiquetas o indicadores arbitrarios de las propiedades que representan.

Los defensores del neo-empirismo rechazan la idea de que los conceptos son representaciones amodales que funcionan como indicadores o etiquetas y defienden, en cambio, que son conjuntos de representaciones perceptivas que funcionan como detectores de categorías (Prinz 2002, 2005). Parte de la defensa del neo-empirismo suele basarse en argumentos de parsimonia, señalando que postular un nivel de representaciones amodales es superfluo y que todos los fenómenos conceptuales se pueden explicar apelando solamente a conceptos vehiculados por representaciones perceptivas. En general, la estrategia consiste en señalar que las teorías amodales deben postular representaciones perceptivas (que son mediadoras entre los conceptos y el mundo) y, en cambio, las teorías que postulan vehículos conceptuales modales, no necesitan proponer, además, representaciones amodales. En ausencia de evidencia decisiva a favor de uno u otro tipo de vehículo, una teoría más parsimoniosa resulta preferible (Barsalou 1999, Pecher 2013, Prinz 2002). A continuación presentaré una reconstrucción más detallada de este argumento, en la versión de Prinz.

Tal como Prinz reconstruye la teoría fodoriana, habría un nivel de indicadores y otro de detectores. Los *indicadores* son “entidades no estructuradas que caen bajo el control nomológico de alguna propiedad” (Prinz 2002, p. 123) y funcionan como etiquetas para señalar propiedades en el mundo. Los *detectores*, en cambio, son mecanismos mediadores entre los indicadores y las propiedades que estos últimos indican. Prinz señala que las teorías defensoras de las representaciones amodales no pueden prescindir de los detectores perceptivos, puesto que estos son necesarios para conectar causalmente los indicadores con las propiedades que representan. Contra los defensores de las representaciones amodales, argumenta que los mecanismos de detección realizan todo el “trabajo” conceptual, de manera que los indicadores serían redundantes y por eso concluye que los conceptos *son* mecanismos

detectores.³⁹ Así, sostiene que los conceptos se deben identificar con los conjuntos de representaciones perceptivas que se usan en la detección de categorías, a los que llama *proxitipos*. Para fortalecer este argumento de parsimonia, Prinz señala dos ventajas adicionales que traería identificar los conceptos con conjuntos de representaciones perceptivas. Primero, los detectores participan directamente en las relaciones que les confieren contenido. Si los conceptos son detectores, “cómo un concepto adquiere su contenido intencional está determinado en parte, por cuestiones intrínsecas a él” (Prinz 2002, p. 124). Segundo, y relacionado con lo anterior, los detectores son menos arbitrarios que los indicadores, puesto que mantienen cierto isomorfismo con aquello que detectan. Así, Prinz sostiene que identificar los conceptos con mecanismos detectores no sólo implica una teoría más económica, sino que también ofrece ventajas explicativas.

Ahora bien, el argumento de parsimonia fortalecido funcionaría si la teoría de proxitipos consiguiera efectivamente explicar todos los fenómenos que se propone explicar, i.e. si se pudiera dar cuenta de todos ellos prescindiendo del nivel de los indicadores. Prinz adopta una lista exhaustiva de fenómenos, que incorpora prácticamente todos los requisitos explicativos que se han propuesto en la literatura sobre conceptos: el contenido (Prinz incluye no sólo el contenido intencional, sino también el contenido cognitivo), el alcance, la composicionalidad, la publicidad, la adquisición y la categorización. En §2 señalaré las ventajas que presenta el neo-empirismo en relación a los requisitos de categorización y adquisición. Sin embargo, mostraré luego, en §3, que hay al menos dos fenómenos que no pueden explicarse sin postular indicadores. Mostraré que las representaciones perceptivas no resultan suficientes para dar cuenta del contenido (§3.1), el alcance (§3.2) ni la composicionalidad (§3.3) y que para explicarlos es necesario incorporar representaciones amodales. En consecuencia, el argumento de parsimonia no resulta suficiente para rechazar las representaciones amodales. Como señalaré hacia el final del capítulo, una propuesta de tipo pluralista que postule tanto vehículos perceptivos como amodales implicaría una ventaja para dar cuenta de los distintos requisitos explicativos.

Con respecto a la estructura semántica de los conceptos, el neo-empirismo propone que los conceptos presentan una estructura interna: son conjuntos de representaciones multimodales, i.e. representaciones que corresponden a distintas modalidades perceptivas. Con respecto a las propiedades semánticas, Prinz propone que los conceptos tienen dos tipos

³⁹ Prinz se refiere a los conceptos tanto en términos de representaciones mentales (e.g. 2002, pp. 108, 149) como mecanismos (e.g. 2002, pp. 125-126, 237). Esto resulta ambiguo, pues las representaciones mentales serían estrictamente *partes de* un mecanismo y no mecanismos en sí.

de contenido: un contenido intencional o real y un contenido cognitivo o nominal. El contenido real refiere a esencias reales (aunque ocultas) de las cosas. El contenido nominal, en cambio, refiere a las propiedades a partir de las cuales identificamos las instancias en el mundo que caen bajo un concepto. Así, ORO, por ejemplo, tiene dos tipos de contenido distintos. Por un lado, tiene un contenido o referente real, la esencia de ORO, el elemento *Au*, con el número atómico 79. Por otro lado, tiene un contenido nominal, el conjunto de propiedades aparentes por medio de las cuales identificamos las instancias de ORO: *amarillo, brillante, maleable* (Prinz, 2002).

Para explicar estos dos tipos de contenido, Prinz adopta una versión de la semántica de doble factor. Propone una versión novedosa, dado que ambos se determinan del mismo modo, por medio de una semántica informacional. Como vimos en el capítulo 4, esta última establece que el contenido está determinado por relaciones nomológicas entre instancias de los conceptos e instancias de las propiedades a las cuales estos refieren, de modo que el contenido de un concepto covaría con las instancias de las propiedades en el mundo que ese concepto expresa. Prinz sostiene que el contenido real covaría con las propiedades esenciales de las categorías, y el contenido nominal, en cambio, covaría con sus propiedades de apariencia. De este modo, tanto el contenido real como el nominal se individualizan de manera externalista, esto es, sus referentes son propiedades en el mundo, con las cuales mantienen relaciones nomológicas. Para determinar el contenido real, además de la condición de covariación nomológica, se agrega una condición etiológica, que alude a la historia causal del concepto. Según esta última condición, los conceptos refieren a sus causas incipientes, i.e. a las propiedades que de hecho causaron las primeras instancias de ese concepto.

Prinz apela a estos dos tipos de contenido para poder individualizar conceptos en los casos Frege y los casos de Tierra Gemela. Como vimos en relación al atomismo informacional, estos casos, especialmente los casos Frege, resultan difíciles de explicar si se atribuye a los conceptos únicamente un contenido real. Así, para poder individualizar los conceptos de manera fina, Prinz propone que estos tienen un contenido real y uno nominal. La articulación entre los dos tipos de contenido no resulta sencilla. Prinz (2002 p. 270) señala que los proximitivos *constituyen* el contenido nominal, puesto que consisten en un conjunto de representaciones perceptivas y las representaciones perceptivas primitivas que los componen se individualizan por las apariencias que detectan. En principio, el contenido nominal “aproxima” el contenido real y, de esta manera, sirve para detectar ese contenido. El contenido nominal sería un intermediario entre los conceptos y sus contenidos reales. En §3 señalaré que la conjunción de esta teoría semántica y el rechazo de un nivel de indicadores o

etiquetas conduce a dificultades relativas a la individuación de conceptos.

En lo que sigue señalaré, en primer lugar, los principales estudios que ofrecen evidencia a favor del neo-empirismo en general. En segundo lugar, señalaré las ventajas de este enfoque en relación a los requisitos de categorización y adquisición. Para esto me concentraré en la teoría de proxitipos de Prinz, dado que ofrece una teoría semántica detallada que me permitirá evaluar la satisfacción del requisito de contenido. En relación a la categorización, por otra parte, la teoría de proxitipos propone articular distintos tipos de estructuras internas, de modo que me permitirá también analizar ciertas tesis pluralistas.

2. La evidencia y las ventajas del neo-empirismo

En esta sección revisaré, en primer lugar, la evidencia empírica que sugiere que las representaciones conceptuales están vehiculadas en un formato de tipo perceptivo. Algunos de estos estudios se proponen para tareas de verificación de propiedades que involucran procesos de categorización, en el sentido de anticipar las características que tendrá un ejemplar por pertenecer a una categoría. Así, estos estudios agregan una serie de efectos que se deberían tener en cuenta al tratar el requisito de categorización (§2.1). Destacaré las ventajas que tiene la teoría de proxitipos de Prinz para dar cuenta de los distintos fenómenos observados en tareas de categorización, dado que no sólo explica los fenómenos asociados a los vehículos perceptivos, sino que al proponer que los conceptos tienen partes con distintos tipos de estructura semántica (como prototipos, ejemplares y teorías), está en condiciones de dar cuenta del resto de los efectos observados en estudios de categorización, presentados en los capítulos 2 y 3 (§2.2). Sostendré que, además, al postular que los conceptos son copias de representaciones perceptivas, cuenta con una ventaja para explicar la adquisición (§2.3).

2.1 La evidencia a favor de los vehículos modales

Distintos estudios que involucran diversos tipos de tareas apoyan la tesis de los vehículos perceptivos. Algunos de los principales se basan en tareas de categorización, puesto que proponen verificar si una propiedad corresponde a las instancias de una determinada categoría, e.g. si dulce es una propiedad de AZÚCAR. Distintos estudios con este tipo de tareas registran un costo en el tiempo de procesamiento asociado al cambio de modalidad entre propiedades sucesivas, i.e. un aumento en el tiempo de reacción para verificar propiedades de un concepto cuando hay un cambio de modalidad con respecto al estímulo anterior, así como un aumento en la tasa de error. Por ejemplo, los participantes tardan más en responder al par LICUADORA-*ruidosa* si éste implica un cambio de modalidad con

respecto al par anterior, i.e. si en el ejemplo inmediatamente anterior se debía evaluar una propiedad de una modalidad diferente, como en el par ARÁNDANOS-*ácidos*, que si se trataba de una propiedad de la misma modalidad, como en HOJAS-*crujientes* (Marques 2006, Pecher *et al.* 2003, 2004). Según los defensores del neo-empirismo, estos resultados se explican porque para verificar las propiedades activamos una simulación del concepto en cuestión, enfocando la atención en la modalidad de la propiedad dada. El aumento en el tiempo de procesamiento asociado al cambio de modalidad se debería al cambio de atención entre los distintos sistemas perceptivos. Por ejemplo, se debe dirigir la atención que estaba centrada en el sistema gustativo, al auditivo. Estos estudios sugieren que activar representaciones conceptuales involucra activar representaciones de los distintos sistemas perceptivos.

Otro tipo de estudio sugiere que las representaciones que codifican orientaciones espaciales guardan cierto isomorfismo con aquello que ocasionaría estas representaciones si se presentara ante la percepción. Stanfield y Zwaan (2001), pusieron a prueba la tesis del isomorfismo a partir de una de sus predicciones: una representación de, por ejemplo, “Juan puso el lápiz en el lapicero” incluirá una orientación vertical del lápiz, mientras que “Juan puso el lápiz en el cajón” incluirá una orientación horizontal del lápiz (Barsalou 1999). En el estudio de Stanfield y Zwaan, los participantes debían determinar si una imagen representaba una acción que había sido previamente descrita en una oración. Las acciones tenían una orientación vertical u horizontal. Este estudio indicó que los tiempos de respuesta eran más breves cuando las imágenes tenían la misma orientación que la acción descrita, lo cual sugiere que tenían una representación perceptiva disponible con la orientación en cuestión.

Otra serie de estudios se proponen determinar el rol de las instrucciones motoras en las representaciones conceptuales. Con este objetivo, Glenberg y Kaschak (2002) elaboraron un diseño experimental en el que los participantes debían responder si determinadas oraciones tenían sentido, y para responder debían realizar un movimiento con la mano acercándola o alejándola del cuerpo. Los autores reportaron un efecto al que denominan *compatibilidad acción-oración*, según el cual resulta más difícil responder cuando el movimiento de la respuesta se dirige en sentido contrario al del movimiento implícito en la oración. Por ejemplo, si para indicar que una oración tiene sentido, los participantes debían responder acercando la mano al cuerpo, les resultaba más fácil evaluar “Abrir el cajón”, que implica un movimiento en la misma dirección, que “Cerrar el cajón”, que involucra un movimiento en la dirección contraria. Estos resultados sugieren que las representaciones (o instrucciones) motoras también están involucradas en la codificación de las representaciones

conceptuales. En general, el neo-empirismo suele entender “representaciones perceptivas” en un sentido amplio, incluyendo todas las producidas por los sistemas sensoriomotores.

Otros tipos de estudios se proponen testear cómo se ve afectado el procesamiento conceptual por la carga en los sistemas de procesamiento perceptivo, con el objetivo de mostrar que ambos procesamientos involucran los mismos sistemas. En este sentido, se han reportado dos tipos de interacción, dependiendo de si la carga en los sistemas perceptivos y conceptuales ocurre de manera alternada o concurrente. Con respecto al primer tipo de interacción, modificando el paradigma de verificación de propiedades, van Dantzig y sus colegas (van Dantzig, Pecher, Zeelenberg & Barsalou 2008) alternaron tareas de detección perceptiva, en las que los participantes debían indicar si un estímulo, que podía ser visual auditivo o táctil (una luz, un tono o una vibración, respectivamente) aparecía a la derecha o a la izquierda de una pantalla, con tareas de verificación de propiedades de categorías, como las descritas más arriba. Los resultados indicaron que, al igual que en los estudios que sólo evalúan la verificación de propiedades, al alternar modalidades se observa un costo en el tiempo de respuesta en comparación con la condición en la que la modalidad es la misma (aunque la tasa de error no varía significativamente en las dos condiciones). Estos resultados sugieren que la activación de representaciones en un sistema perceptivo en tareas puramente perceptivas, en las que los estímulos no implican referencia a nada concreto, afectan el procesamiento conceptual. Los autores explican esta interacción como un efecto de *priming*, i.e. la activación de, e.g. una luz, facilita la evaluación posterior de “el brócoli es verde”, en comparación con la evaluación de “las abejas zumban”.

Con respecto al segundo tipo de interacción, en estudios que, en lugar de presentar tareas que implican una carga alternada en los sistemas perceptivos (i.e. tareas sucesivas, ya sea ambas de verificación de propiedades o primero de detección perceptiva y luego de verificación de propiedades), presentan tareas concurrentes, se observa el efecto contrario. Es los casos en los que se produce una carga en la memoria de trabajo de manera simultánea con tareas de verificación de propiedades, la tarea se dificulta más cuando ambas involucran la misma modalidad perceptiva, que cuando involucran modalidades diferentes. Vermeulen y sus colegas instruyeron a los participantes de un estudio para que recordaran una serie de tres tonos o imágenes (puntos de colores o figuras que no representaban nada concreto), para reconocerlos más adelante. Por ejemplo, los participantes observaban tres imágenes que debían recordar *durante* una tarea de verificación de una propiedad para luego, ante una serie de tres imágenes, indicar si eran idénticas o diferentes a las primeras. Los resultados muestran que cuando la modalidad en las dos tareas es la misma, tanto la verificación de la propiedad

como la recolección de los estímulos que debían recordar se ven dificultadas. Cuando la modalidad es distinta, en cambio, los participantes no exhibieron dificultades para llevar a cabo la verificación de propiedades ni la recolección de aquello que debían recordar (Vermeulen, Corneille & Niedenthal 2008).⁴⁰

Así, esta serie de resultados empíricos sugiere que pensar involucra activar los sistemas perceptivos.⁴¹ Una de las objeciones principales a la evidencia que se suele presentar a favor del neo-empirismo, señala que los mismos efectos podrían explicarse apelando a una organización de las representaciones amodales por modalidades, valga la aparente contradicción. Esto es, podría haber asociaciones entre los símbolos amodales que refieren a propiedades de cada modalidad, de modo que la activación de una representación que refiere a una propiedad, por ejemplo, visual, facilite la activación de otras representaciones de propiedades visuales. También se podría postular, alternativamente, que la información conceptual está organizada espacialmente en el cerebro por modalidades. Como afirma Machery (2006), una teoría amodal podría sostener que la instanciación cerebral de los conceptos que corresponden a propiedades perceptivas está ubicada cerca de los sistemas perceptivos de esa modalidad. Por ejemplo, AMARILLO estaría localizado cerca del sistema visual y AGUDO estaría ubicado cerca del sistema auditivo. De este modo, se podría explicar el fenómeno observado en los estudios de Pecher y sus colegas, que sugiere un efecto de *priming* entre activaciones de conceptos de la misma modalidad, porque los conceptos de una misma modalidad estarían más cerca uno del otro que de conceptos de otras modalidades.

Los defensores del neo-empirismo reconocen que se pueden ofrecer explicaciones alternativas para la evidencia reportada en los estudios de verificación de propiedades. Por este motivo, apelan a otras características de las teorías, como la parsimonia, para fortalecer sus defensas, o señalan que las teorías amodales deben incorporar hipótesis *post hoc* para explicar los fenómenos en cuestión, mientras que el neo-empirismo los predice. Sin embargo, más allá de estos argumentos y las consideraciones de parsimonia, también se puede señalar que los dos últimos estudios que mencioné, que diferencian entre los efectos de la carga de procesamiento perceptiva alternante y la carga concurrente con tareas conceptuales ofrecen evidencia más fuerte a favor del uso de las representaciones perceptivas como vehículos para las tareas conceptuales. En particular los resultados obtenidos por Vermeulen y sus colegas

⁴⁰ Por otra parte, también se registró una relación con la medida en que se carga la memoria de trabajo, puesto que cuando la carga en la memoria de trabajo era reducida, i.e. en los casos en que se limitaba a un solo ítem, no se observaron efectos en las tareas siguientes.

⁴¹ Los estudios que presenté aquí son ejemplos de los que considero que ofrecen evidencia más significativa a favor del neo-empirismo. Para una presentación más abarcativa de los distintos estudios elaborados por los defensores del enfoque en las últimas décadas, véase Pecher (2013).

sugieren que cuando la carga es concurrente se produce una competencia por el uso de los mismos recursos, i.e. por los mecanismos de detección perceptiva, lo que implicaría una participación constitutiva de estos mecanismos en los vehículos conceptuales, que resulta más difícil de explicar por medio de sistemas de símbolos amodales.

Ahora bien, esta evidencia no resulta, sin embargo, decisiva para afirmar que *todos* los conceptos están vehiculados por representaciones perceptivas y que no disponemos, además, de representaciones amodales. En §3 señalaré las dificultades que acarrea el rechazo de las representaciones amodales. Antes de ocuparme de tales dificultades, destacaré las ventajas del neo-empirismo para dar cuenta de los requisitos de categorización y adquisición.

2.2 Categorización

Además de explicar los distintos efectos revisados en los últimos párrafos, el neo-empirismo, particularmente en la versión de la teoría de proxitipos, está en una mejor posición para dar cuenta de la categorización que las teorías correspondientes a los enfoques analizados en los capítulos 2, 3 y 4. Como vimos, el atomismo informacional no está en condiciones de dar cuenta de los efectos observados en tareas de categorización, puesto que rechaza cualquier tipo de estructura interna. Las teorías de prototipos, ejemplares y teorías, por otra parte, resultan adecuadas para explicar algunos de los efectos observados, pero ninguna consigue acomodarlos a todos. La teoría de proxitipos cuenta con la ventaja de incorporar distintos tipos de estructuras semánticas, todas representadas como conjuntos de rasgos, lo que le proporciona las herramientas representacionales para dar cuenta de los distintos efectos observados.

Para explicar la categorización, Prinz propone que ésta puede involucrar procesos de distinto tipo, en función del tipo de información que se utilice. Según la teoría de proxitipos, los conceptos son construcciones altamente variables, en el sentido que la representación que se activa en un contexto no será la misma que se active en otro. Prinz (2002) explica esto en relación al *uso* de los conceptos. Puesto que los conceptos se usan como mecanismos para detectar instancias de categorías, la representación que se activa en cada caso está relacionada con el contexto. No será la misma representación la que usemos para rastrear un perro guardián o un perro en el ártico, aunque ambas son proxitipos del concepto PERRO.⁴² De acuerdo con esta propuesta, las representaciones perceptivas se almacenan en redes de memoria a largo plazo y cada concepto puede tener diferentes proxitipos en la memoria de

⁴² Esta tesis acarrea dificultades en relación a la individuación del contenido de los conceptos, de las que me ocuparé en la próxima sección.

trabajo según la información que se active.

Pues bien, en los procesos de categorización, el contexto de la tarea determinará la activación de uno u otro tipo de información. En algunos casos lo que sucede es que representamos la instancia que queremos categorizar (ya sea una instancia que percibimos directamente o que escuchamos describir o nombrar) por medio de una serie de rasgos. Esta serie de rasgos se compara con nuestras redes de memoria a largo plazo. El proceso de comparación se puede modelar por medio de una función que arroja una medida de la similitud entre ambos conjuntos de rasgos. Al igual que sugiere el enfoque de ejemplares, categorizamos la instancia en cuestión como correspondiente a la categoría de la red que contenga la representación más similar. Esto permite explicar los efectos de familiaridad con instancias conocidas, de los que me ocupé en el capítulo 2. Alternativamente, podemos determinar que una instancia pertenece a una categoría por su similitud con las tendencias centrales de la categoría, almacenadas en una representación prototípica, como sugiere el enfoque de prototipos. Este tipo de proceso, a su vez, permite dar cuenta de los efectos de tipicidad, también vistos en el capítulo 2.

Por otro lado, para dar cuenta de los efectos observados en los ejemplos de transformación de Keil, por ejemplo, Prinz (2002) sostiene que en esos casos, los rasgos a partir de los cuales se construye la instancia que se quiere categorizar están asociados a la red de memoria a largo plazo correspondiente a la categoría con la que se los describe inicialmente (recordemos que en los ejemplos se describe inicialmente un miembro de una categoría, e.g. un caballo, al que luego se modifican sus rasgos superficiales para que parezca una cebra. En este ejemplo, los rasgos a partir de los que se representa la instancia estarían ligados a la red de CABALLO, la cual incluye el conocimiento de que hay una esencia oculta que determina la pertenencia a la categoría. Así, al modificarse los rasgos superficiales de esa instancia, la asociación con la categoría inicial no se perdería. De este modo, la teoría de proxitipos también puede dar cuenta de los efectos de los que se ha ocupado el enfoque de la teoría-teoría, que vimos en el capítulo 3. Dado que incorpora distintos tipos de estructuras de información, la teoría de proxitipos puede dar cuenta de los principales efectos observados en tareas de categorización, lo cual le otorga una ventaja frente a los enfoques que incorporan un único tipo de estructura.

2.3 Adquisición

Una de las ventajas principales del neo-empirismo es poder explicar la adquisición de conceptos como resultado de la experiencia, a partir de las interacciones perceptivas con el

entorno. Esta teoría retoma del empirismo clásico la tesis de la prioridad perceptiva y sostiene que los conceptos tienen un origen perceptivo. De esta manera establece una continuidad entre la percepción y la cognición. La prioridad de la percepción con respecto a la cognición es de tipo causal y establece que “los conceptos no llegarían a existir, si no hubiera representaciones perceptivas disponibles para ser copiadas o combinadas” (Prinz 2002, p. 108). El hecho de que los conceptos sean copias de representaciones perceptivas que se almacenan en redes de memoria a largo plazo y pueden ser reactivadas en la memoria de trabajo, podría sugerir que adquirir representaciones perceptivas equivale a adquirir conceptos. Sin embargo, Prinz introduce una distinción entre las meras representaciones perceptivas y las representaciones perceptivas conceptuales (los conceptos). Los conceptos son representaciones perceptivas (o copias de ellas) que están bajo control endógeno del organismo, i.e. que pueden ser activadas con independencia de los estímulos perceptivos que las hubieran ocasionado.

Así, se ha adquirido un concepto cuando se tiene la capacidad para reactivar de las representaciones que conforman ese concepto de manera independiente del estímulo. Las representaciones perceptivas están bajo control del entorno en el sentido en que dependen de que se esté en contacto con determinados estímulos para que se activen. Los conceptos, en cambio, se pueden activar por control endógeno, esto es, en ausencia de los estímulos relevantes en el entorno. Podemos, por ejemplo, desplegar o reactivar las representaciones perceptivas que necesitamos para planear un viaje, sin necesidad de tener en el entorno aviones, hoteles ni valijas. De este modo, puede decirse que la adquisición de conceptos para el neo-empirismo es un pasaje de representaciones perceptivas pre-conceptuales a representaciones perceptivas conceptuales, y en este sentido, el formato de representación en ambos casos es el mismo.

Integrando las distintas consideraciones que ofrece Prinz, la adquisición de conceptos puede reconstruirse en tres etapas. En la primera etapa se produce la *activación de representaciones perceptivas*, i.e. la formación de representaciones perceptivas a partir de la interacción con el entorno. En una segunda etapa se produce la *adquisición de conceptos*, por la cual las representaciones pasan a estar bajo control endógeno y, por último, también puede producirse una tercera etapa de *aprendizaje*, por la cual los conceptos (o representaciones perceptivas conceptuales) se van modificando con el tiempo a partir del contacto con instancias de las categorías y con nueva información. Así, según esta teoría, los encuentros con instancias de categorías causan que se activen representaciones perceptivas en nuestros sistemas de *input* dedicados y hay un mecanismo que causa que se almacenen copias de estas

representaciones para ser reactivadas en ocasiones posteriores. Por otra parte, una vez que se van adquiriendo conceptos, operan los mecanismos de aprendizaje. Se pueden aprender características de los mismos almacenando otras representaciones asociadas e ir “afinando” las categorías ajustando los pesos de los distintos rasgos en las representaciones prototípicas.

Barsalou (1999, Barsalou *et al.* 2003) también explica la adquisición de conceptos como el registro de la activación de representaciones en los distintos sistemas perceptivos, que posteriormente pueden reactivarse en ausencia de los estímulos. Al percibir instancias de una categoría, se activa una serie de detectores de rasgos en cada uno de los sistemas perceptivos. Así, por ejemplo, en interacciones con perros se pueden producir representaciones en el sistema visual, el auditivo, el olfativo y el motor. Las interacciones frecuentes con los miembros de la categoría dan lugar a registros de la activación en cada uno de los sistemas, que son almacenados en la memoria. Estos registros permiten reactivar esas representaciones con posterioridad para pensar acerca de perros. Se trata de registros multimodales, puesto que incluyen asociaciones entre las representaciones de distintas modalidades que fueron activadas frente a sus instancias.

Barsalou se refiere a los procesos posteriores de reactivación de representaciones perceptivas como procesos de simulación, puesto que se simulan los estados perceptivos que serían causados por instancias de la categoría en la que se piensa. Las simulaciones no reproducen exactamente los estados perceptivos originales, puesto que los registros de las primeras activaciones no almacenan todos los detalles de los estados perceptivos, sino que involucran cierto grado de abstracción y pueden estar influidos por sesgos atencionales. Por otra parte, a medida que acumulamos experiencias con instancias de una categoría, los registros se van modificando, alterando la fuerza de las asociaciones entre las distintas representaciones perceptivas, e incorporando representaciones nuevas.

Así, el enfoque neo-empirista, en sus distintas versiones, tiene la ventaja de explicar la adquisición de conceptos como una consecuencia de la interacción con instancias de la categoría y, con ello, la formación de representaciones en los sistemas perceptivos. Sin embargo, es poco plausible que esta explicación pueda extenderse a todos los conceptos que poseemos, puesto que, como argumentaré en la próxima sección, en relación al requisito de alcance, los conceptos abstractos y teóricos, no pueden adquirirse por encuentros perceptivos con sus instancias.

3. Las dificultades del neo-empirismo

En lo que sigue señalaré las principales dificultades del neo-empirismo en relación a

los requisitos de contenido, alcance y composicionalidad. Intentaré mostrar que estas dificultades son propias de los vehículos perceptivos y se deben al intento de dar cuenta de todos los fenómenos sin recurrir a representaciones amodales. En este sentido, en el mismo espíritu pluralista de los capítulos anteriores, intento mostrar que, si bien los vehículos perceptivos presentan ciertas ventajas, adherir a la tesis de la uniformidad con respecto a los vehículos acarrea limitaciones explicativas y que deben postularse, además, representaciones amodales.

3.1 Intencionalidad

Una primera dificultad se relaciona con el requisito de intencionalidad. Presentaré dos argumentos con el objetivo de mostrar que la teoría de proxitipos no consigue dar cuenta de la individuación de conceptos sin tener que recurrir al nivel de los indicadores. El primer argumento tiene dos partes. La primera señala que si se identifican los conceptos con proxitipos y estos se individuán por las propiedades de apariencia que detectan, entonces los conceptos quedarán reducidos a un puro contenido nominal, perdiendo su referencia a propiedades esenciales.

Prinz (2002, p. 124) afirma que las representaciones que permiten detectar categorías, i.e. los proxitipos, intervienen en las relaciones que determinan el contenido. Así, puesto que las representaciones a partir de las cuales detectamos las categorías constituyen el contenido nominal, el contenido real no dependería sólo de relaciones nomológicas con propiedades esenciales, sino que estaría determinado, en parte, por el contenido nominal. Pero si esto es así, se introduce el riesgo de que los conceptos tengan sólo contenido nominal y pierdan su contenido real. En la segunda parte del argumento discutiré la posibilidad de individuar los conceptos por su contenido real. Si esto fuera posible, se podría eludir la primera crítica. Sin embargo, argumentaré que la individuación de conceptos en tanto tipos reales no es posible recurriendo únicamente a representaciones perceptivas. En el segundo argumento presento un caso límite en el cual el propio Prinz se vería forzado a introducir representaciones en un nivel distinto al de los detectores. Los dos argumentos intentan mostrar que este enfoque no logra dar cuenta del fenómeno del contenido sin apelar a representaciones que no sean perceptivas. Sostendré que para resolver los problemas de individuación del contenido de los conceptos es necesario apelar al nivel de los indicadores, i.e. postular un nivel de representaciones amodales.

Prinz afirma, como reconstruí en 1, que los conceptos son proxitipos y define a estos últimos como conjuntos de representaciones perceptivas que se activan en la memoria de

trabajo para detectar categorías. Estos conjuntos varían de acuerdo al contexto. Como mencioné, no utilizamos el mismo conjunto de representaciones para rastrear un perro guardián que para rastrear un perro en el Ártico. Ahora bien, cabe preguntarse si el proxitipo que se activa para detectar un perro guardián y el que se activa para detectar un perro en el Ártico son distintos proxitipos correspondientes al mismo concepto PERRO, o se trata de dos conceptos distintos. El problema de la individuación de conceptos en esta teoría radica en cómo entender la relación entre conceptos y proxitipos. Prinz sostiene que distintos proxitipos pueden instanciar *el mismo* concepto en distintos contextos:

Los conceptos pueden representar categorías aun si son instanciados de maneras distintas. Desde la perspectiva que estoy considerando, el mismo concepto estará constituido por representaciones diferentes en distintas ocasiones. (2005, pp. 685-686)

De esta manera, no tenemos un concepto diferente cada vez que pensamos acerca de perros, sino que usamos distintos proxitipos correspondientes al mismo concepto. Así, los conceptos no pueden individuarse por el conjunto particular de representaciones perceptivas que los instancia en cada ocasión, puesto que de ese modo, cada instancia sería un concepto distinto. Si distintos pensamientos, tales como que un perro guardián está en un jardín y que un perro está en el Ártico, están constituidos por el mismo concepto PERRO, a pesar de estar compuestos por proxitipos distintos, se sigue que los conceptos no pueden ser idénticos a proxitipos y que no pueden individuarse por estos últimos (que serían sus contenidos nominales), sino más bien por sus contenidos reales. Pero el contenido real no resulta suficiente para la individuación, puesto que no permite dar cuenta de los casos Frege, de modo que esto parece señalar nuevamente al contenido nominal como el responsable último de la individuación de conceptos. Sin embargo, el contenido nominal equivale a apariencias. Si los conceptos se individuaran por sus apariencias, no se podría dar cuenta de los casos de Tierra Gemela ni de los casos de gemelos en esta tierra, i.e. los casos en los que dos conceptos tienen distinto contenido real (por ejemplo, tienen por referentes sustancias distintas, tales como las famosas H₂O y XYZ), a pesar de que sus referentes tienen las mismas propiedades de apariencia.

Una solución posible es apelar a la individuación por tipos, de modo que distintos casos de proxitipos sean idénticos en tipo. Así, el problema de la individuación se resolvería si hubiera un modo de individuar los conceptos en tanto tipos y un modo de establecer que

distintos conjuntos de representaciones perceptivas (distintos proxitipos) son instancias del mismo concepto tipo. Esta es la estrategia que adopta Prinz (2002, 2003), de manera que debería especificar cuándo dos conjuntos de representaciones perceptivas en la memoria de trabajo corresponderían al mismo tipo de concepto. Pero esto no parece posible, puesto que la dualidad se traslada a los conceptos en tanto tipos, debido a que sólo es posible individuar tipos reales y tipos nominales de manera independiente, como desarrollaré a continuación.

Cada conjunto de rasgos perceptivos (o con base perceptiva) que nos permite detectar una categoría al activarse en la memoria de trabajo constituye un proxitipo. Y cada uno de esos conjuntos de rasgos conforma un contenido nominal. En este sentido, cada proxitipo *constituye* un contenido nominal, de manera que “en tanto tipos nominales, los conceptos son proxitipos” (Prinz 2002, p. 279). Ahora bien, los conceptos no pueden identificarse únicamente con el contenido nominal, por lo que la identidad entre proxitipos y conceptos no puede mantenerse cuando se considera a los conceptos como tipos reales: distintos proxitipos pueden corresponder al mismo contenido real y distintos contenidos reales pueden corresponder al mismo proxitipo (Prinz 2002, p. 279). El problema de la individuación evidencia el problema de la relación entre conceptos y proxitipos, en tanto muestra que las nociones de “proxitipo” y “concepto” no son las mismas, puesto que no hay una correspondencia uno a uno entre ellas. Los conceptos en tanto tipos nominales pueden ser entendidos como idénticos a los proxitipos, pero los conceptos en tanto tipos reales no, sino que son conjuntos de proxitipos distintos que comparten el mismo contenido real. En consecuencia, afirmar la identidad entre conceptos y proxitipos implicaría reducir los conceptos a su contenido nominal, excluyendo el contenido real, y esto sería, para Prinz, ir en contra de su propia semántica de conceptos.⁴³

Por otra parte, es cuestionable que Prinz pueda dar cuenta de la individuación de proxitipos en tanto tipos reales. Las copias de representaciones perceptivas que usamos para detectar instancias de categorías se almacenan en la memoria a largo plazo formando redes de representaciones ligadas entre sí, de modo que la información en un proxitipo está articulada a partir de distintas asociaciones entre representaciones. Prinz no desarrolla en profundidad los principios a partir de los cuales se producen estas asociaciones, pero menciona dos: la co-instanciación de representaciones en un mismo objeto y la similitud entre representaciones. El primero da lugar a enlaces *jerárquicos*, que permiten “acercar” o “alejarse” las representaciones (focalizando partes específicas), enlaces *transformacionales*, que agrupan

⁴³ De Rosa (2005) llega a una conclusión análoga a partir de otras consideraciones en torno a la relación entre proxitipos y conceptos.

transformaciones posibles de una representación en otra (por ejemplo, un animal en distintas posiciones), enlaces *de unión*, que asocian representaciones de distintas modalidades, y enlaces *situacionales*, que agrupan representaciones de distintos objetos que generalmente co-ocurren (por ejemplo, de un gorila y una banana). El segundo principio da lugar a los enlaces *predicativos*, i.e. la disposición a transferir características de una representación a otra (Prinz 2002, p. 145).

Prinz delega la individuación de conceptos a los enlaces *predicativos* que se establecen en la memoria a largo plazo e identifica dos representaciones como manifestaciones del mismo concepto “si están predicativamente ligadas a la misma red de memoria” (Prinz 2003, p. 298). Si la individuación de conceptos depende de la asociación predicativa con una red en la memoria a largo plazo, es necesario, por un lado, poder individuar las redes y, por otro lado, determinar de qué red deriva cada proxitipo. Ambas tareas resultan difíciles si sólo se dispone de representaciones perceptivas asociadas. Tal como señala Ryder (2003), la superposición entre redes, debido a las múltiples asociaciones entre representaciones genera dudas respecto de la posibilidad de identificación de una red correspondiente a un único concepto. Los enlaces que Prinz propone, a partir de los principios de coinstanciación y similitud, no parecen suficientes para individuar redes de representaciones ligadas entre sí.

Una adaptación del ejemplo de Ryder permite ilustrar esta crítica. Supongamos que activo en mi memoria de trabajo un proxitipo cuyo contenido nominal es una apariencia de Bulldog y su contenido real es *ser perro*. Este proxitipo puede estar asociado en la memoria a largo plazo con múltiples representaciones, por ejemplo, PULGA, PLAZA, VETERINARIO y muchos otros proxitipos (además de los de las otras razas de perros). Prinz afirmaría que el proxitipo de apariencia de Bulldog deriva del concepto PERRO y no del concepto PULGA. Según Prinz (2002), esto es así porque la representación de Bulldog está predicativamente ligada a PERRO, mientras que la relación entre BULLDOG y PULGA se daría mediante otro tipo de enlace. Sin embargo, se puede objetar, siguiendo a Ryder, que Prinz está apelando a una diferencia entre enlaces intra-redes (i.e. entre PERRO y BULLDOG) e inter-redes (i.e. entre BULLDOG y PULGA), pero la definición de enlace predicativo como la disposición a transferir características de una representación a otra, no resulta suficiente para una distinción tan fina. ¿Por qué esta disposición se daría únicamente entre representaciones de la misma categoría? Esto no está justificado en la teoría.

Además, los enlaces predicativos se basan en la similitud perceptiva, pero es cuestionable qué características similares tienen, por ejemplo, un chihuahua y un gran danés,

además de ser ambos precisamente perros. La disposición a transferir propiedades de CHIHUAHUA a GRAN DANÉS no está justificada perceptivamente. No transferiríamos de uno a otro propiedades de color, ni de tamaño, y por otro lado también estaríamos dispuestos a transferir la característica de tener cuatro patas a CABALLO, RATA o HIPOPÓTAMO (Ryder 2003). Respecto de este problema, Sarnecki (2004) señala que los enlaces predicativos no parecen tener una base puramente perceptiva, con lo cual no sólo no son suficientes para dar cuenta de la individuación de redes, sino que también es difícil fundamentar las relaciones predicativas mismas recurriendo únicamente a representaciones perceptivas. Las dificultades para distinguir entre enlaces inter-redes e intra-redes parecen apuntar a la necesidad de recurrir a algún tipo de indicador para individuar conceptos. Frente a las dificultades vistas para la individuación, resulta tentador apelar a cierta organización de las representaciones en archivos mentales en los cuales las representaciones perceptivas se agrupen para constituir conceptos. A continuación presentaré un caso en el que Prinz introduce estos archivos y argumentaré que estos deben incorporar igualmente símbolos amodales.

Prinz (2002) utiliza la metáfora de las carpetas de archivos cuando considera la posibilidad de que un mismo contenido nominal corresponda a dos conceptos diferentes, sin que haya ninguna representación perceptiva asociada a alguno de ellos que permita distinguirlos. El ejemplo que propone es el de una persona que cree que hay dos individuos distintos llamados Farrakhan, pero no tiene ninguna información asociada a ninguno de los dos. La explicación propuesta para este caso asigna un rol fundamental a los archivos mentales en los cuales se almacenarían las representaciones: “Una persona en esta situación tiene representaciones distintas en la memoria a largo plazo (piense en dos carpetas mentales sin contenido), pero sus contenidos cognitivos son idénticos” (Prinz, 2002, p. 271). En este caso, los contenidos cognitivos son idénticos, puesto que no hay representaciones perceptivas distintas asociadas a ninguno de los dos conceptos. Por lo tanto, lo único que permitiría distinguirlos es su pertenencia a dos “carpetas” distintas.

En este ejemplo, Prinz sostiene que hay conceptos distintos almacenados en distintos archivos. Considero, por lo tanto, que está recurriendo a un nivel distinto del de los detectores, tal como el de los indicadores que proponen los defensores de las representaciones amodales. Prinz (2004, p. 428) argumenta que los archivos mentales no deben reificarse, sino que “Los ítems en la base de datos se agrupan porque están asociados unos con otros, no con una extraña carpeta color manila en la mente”. Sin embargo, en el caso límite de FARRAKHAN₁ y FARRAKHAN₂ esta estrategia no funciona. En este caso se evidencia la

reificación de las carpetas, porque *no hay* distintos conjuntos de representaciones perceptivas asociados en redes distintas, sino que lo único disponible para distinguir a los dos conceptos son carpetas metafóricas. Si no hubiera absolutamente nada en la mente que instanciara esas carpetas, como los indicadores amodales, la distinción entre esos conceptos se tornaría metafórica también, y en este sentido no habría una distinción de hecho entre ambos. El caso FARRAKHAN muestra que en casos límite en los que no hay información perceptiva es necesario un indicador amodal, i.e. una etiqueta, que distinga conceptos.

En síntesis, si los conceptos son idénticos a proxitipos y los proxitipos constituyen el contenido nominal, entonces los conceptos pierden su contenido real. Si, en lugar de proponer una identidad uno a uno entre conceptos y proxitipos, se propone que cada concepto puede instanciarse en un conjunto de proxitipos distintos, esto impone la necesidad de individuar conjuntos de proxitipos como correspondientes al mismo concepto. Pero esto acarrea el problema de la individuación de redes de representaciones asociadas, que evidencia la necesidad de indicadores amodales. Por otro lado, el ejemplo de FARRAKHAN₁ y FARRAKHAN₂, al igual que las dificultades para distinguir entre enlaces intra-redes e inter-redes, aporta razones para sostener que la organización del nivel de memoria a largo plazo en archivos mentales resulta necesaria para explicar la individuación de conceptos. Para que estos archivos puedan hacer el trabajo de la individuación es necesario que incorporen algo distinto de las representaciones perceptivas asociadas, puesto que, de otro modo, la individuación de archivos sólo duplicaría el problema de la individuación de redes. Lo que permitiría resolver este problema sería incorporar un símbolo amodal a modo de etiqueta. La individuación de redes de representaciones perceptivas asociadas, al igual que el caso de FARRAKHAN₁ y FARRAKHAN₂, requiere un nivel de representaciones distinto del de los detectores, esto es, requiere indicadores amodales.

3.2 Alcance

El análisis del requisito de alcance también revela la necesidad de representaciones amodales. Con respecto a este requisito, este enfoque enfrenta dificultades similares al enfoque de ejemplares, en tanto no consigue dar cuenta de los conceptos abstractos y teóricos. Para el enfoque de ejemplares, la dificultad radica en que no tenemos encuentros directos con instancias de, e.g. electrones. De manera similar, los conceptos abstractos y teóricos no parecen ser reducibles a copias de representaciones perceptivas. Prinz (2002) propone diversas estrategias que permitirían a las representaciones perceptivas representar este tipo de conceptos. Discutiré la aplicación de estas estrategias en dos ejemplos y

argumentaré que las representaciones perceptivas no permiten dar cuenta de la representación de conceptos teóricos y abstractos. Mostraré así que para satisfacer el requisito de alcance resulta necesario incorporar representaciones amodales.

Las estrategias empiristas para representar conceptos teóricos y abstractos pueden clasificarse en dos grupos: las que contribuyen a explicar únicamente el contenido cognitivo de los conceptos y las que explican su contenido real. Entre las primeras, Prinz señala, por un lado, el uso de ciertas habilidades verbales y, por otro, siguiendo a Lakoff y Johnson (1980), la posibilidad de que los conceptos abstractos sean proyecciones metafóricas a partir de representaciones perceptivas básicas relacionadas directamente con la experiencia. Entre las habilidades verbales se cuentan, por ejemplo, saber en qué contexto usar cada palabra o con qué otras palabras está asociada una palabra en particular. Estas habilidades no son suficientes para establecer una relación entre los conceptos y sus referentes en el mundo, porque únicamente reflejan relaciones entre palabras. Por eso no contribuyen al contenido real, sino únicamente al contenido cognitivo. En ese sentido, Prinz considera que nos permiten razonar acerca de propiedades de otro modo intangibles. Por otra parte, la proyección metafórica de representaciones perceptivas, como Prinz mismo señala, excluye un componente central de los conceptos: su contenido real. Según esta estrategia, se establece una relación de semejanza entre las representaciones perceptivas y el concepto en cuestión, pero esta relación no es de equivalencia. De esta manera, las proyecciones metafóricas aproximan el contenido real, pero no lo constituyen. Así, esta estrategia, al igual que la apelación a las habilidades verbales, alude al contenido cognitivo de este tipo de conceptos, pero no da cuenta de cómo estos se relacionan con el mundo, i.e. no da cuenta del contenido real.

Para aplicar la semántica informacional a los conceptos abstractos y teóricos es necesario apelar a una estrategia que los conecte con sus referentes reales. De modo que para no reducirlos a un puro contenido cognitivo es necesario que las representaciones que constituyen los conceptos, i.e. las representaciones que funcionan como mecanismos de detección, estén causalmente conectadas con las propiedades que representan. Para establecer esta conexión, Prinz señala una tercera estrategia: ciertos conceptos están representados por características observables que están asociadas contingentemente a las categorías que representan y que sirven como “rastreadoras de signos”. Habría distintos tipos de rastreo de signos, tales como la identificación de los rasgos superficiales o apariencias de los miembros de una categoría, el uso de instrumentos científicos que sirven como herramientas mediadoras en la percepción, o el rastreo por el uso de palabras del lenguaje natural. Dado que esta es la

única estrategia que explica la conexión entre los conceptos abstractos o teóricos y las propiedades en el mundo a las que refieren, si esta estrategia fallara, este tipo de conceptos quedaría reducido a un puro contenido cognitivo. Como última estrategia, Prinz también sugiere que ciertos supuestos contraejemplos, como NEGACIÓN o DISYUNCIÓN, no corresponden verdaderamente a conceptos, sino a reglas u operaciones. La negación, por ejemplo, sería una variación de una operación de contraste entre las representaciones y el mundo, “una manera especial de medir similitudes entre conjuntos de rasgos” (Prinz 2002, p. 183).⁴⁴

Considero que estas estrategias no son suficientes para explicar cómo se representan los conceptos teóricos y abstractos por medio de representaciones perceptivas. Intentaré mostrar esto tomando como ejemplos a ELECTRÓN y VERDAD. Para dar cuenta del concepto ELECTRÓN, Prinz divide a las mentes que tienen representaciones de electrones entre mentes científicas y mentes no científicas. En el caso de los científicos, Prinz opta por la estrategia del “rastreo de signos” y sugiere que este concepto se puede representar por medio de representaciones perceptivas tales como las de trazos de niebla observables en experimentos científicos al hacer pasar electrones por una cámara que contiene vapor de un gas.⁴⁵ En cambio, el concepto ELECTRÓN que posee la comunidad no científica que domina el uso de la palabra “electrón” aun desconociendo los experimentos con cámaras de niebla, se explica a partir de las habilidades verbales de dicha comunidad. De esta manera, este concepto es deferente, tiene únicamente contenido cognitivo y refiere en virtud de que hay expertos (los científicos) encargados de “anclar” el contenido del concepto en el mundo.

Ahora bien, esta explicación presenta dos inconvenientes. En primer lugar, para toda la población no científica que nunca oyó hablar de cámaras de niebla, la representación del concepto ELECTRÓN no sería más que la de la palabra “electrón”. En ausencia de otros rasgos perceptivos, ELECTRÓN no tendría por contenido cognitivo más que las asociaciones con otras palabras implicadas en el dominio verbal de la palabra “electrón” y no tendría contenido real en absoluto. En segundo lugar, si en las mentes de los científicos ELECTRÓN se representa por trazos de niebla, Prinz no explica cómo puede representarse

⁴⁴ No me ocuparé de este tipo de conceptos. Sin embargo, quiero señalar que Prinz no los trata de manera uniforme, sino que por un lado sugiere que la negación es una operación pero, al mismo tiempo, hace referencia a NO GATO como una representación que tiene un prototipo, i.e. que está estructurada a partir de los rasgos típicos de la categoría (Prinz 2002, p. 287). Si NO fuera sólo un mecanismo especial para medir similitudes, no tendría sentido hacer referencia a un conjunto de rasgos típicos de NO GATO (Barberis 2008).

⁴⁵ La cámara de niebla contiene vapor de un gas saturado, que se condensa ante perturbaciones mínimas, tales como el aumento de presión o temperatura. Al introducir en la cámara partículas con energía, el gas se ioniza y se condensa alrededor de los núcleos que interactuaron con las partículas. Por esta razón se puede observar un camino o trazo de vapor condensado con una forma característica.

perceptivamente el vínculo entre los trazos de niebla y los electrones, dado que no es un vínculo de identidad sino de causa-efecto. Y puesto que el concepto mismo de CAUSACIÓN presenta sus propios problemas, es poco plausible que pueda incorporarse con ventajas explicativas al conjunto de representaciones perceptivas correspondientes a ELECTRÓN.

Respecto del primer inconveniente, el concepto deferente ELECTRÓN no se adquiere por entrar en contacto con la propiedad a la que refiere, sino por medio de representaciones “cuyos significados han sido previamente fijados por otros” (Prinz 2002, p. 252). El contenido real de los conceptos deferentes consiste, de esta manera, en el contenido de las representaciones a partir de las cuales fueron adquiridos. En el caso de ELECTRÓN, la mayoría de la gente entra en contacto con este concepto a partir de una definición como “partícula subatómica cargada negativamente”. Sin embargo, Prinz no pretende sostener un retorno al definicionismo. El contenido real de ELECTRÓN no está constituido por una definición, sino que está fijado por los métodos que los científicos utilizan para reconocer apariencias de electrones. Específicamente, los científicos representan perceptivamente los electrones a partir de los trazos que dejan en cámaras de niebla. De este modo, Prinz resuelve aparentemente la primera objeción, puesto que parece evitar el encierro en un círculo de definiciones y relaciones entre palabras, ofreciendo una explicación de la adquisición del contenido real y cognitivo de conceptos como ELECTRÓN. Sin embargo, esta estrategia supone que hay un mecanismo adecuado que conecta las palabras del lenguaje natural con sus referentes y, de esta manera, no resuelve el problema, sino que lo desplaza hacia una teoría de la referencia del lenguaje natural.

Además de adoptar, tal como Prinz supone, una teoría semántica del lenguaje natural que remite a los expertos científicos para fijar causalmente la referencia de términos como “electrón”, también hay que suponer que los expertos conectan de manera adecuada su propio concepto ELECTRÓN con su contenido real. Como señalo respecto del segundo inconveniente, esto resulta difícil de cumplir si el concepto de los expertos científicos está constituido únicamente por representaciones perceptivas. En el caso de los científicos que representan ELECTRÓN por medio de las representaciones de los trazos en una cámara de niebla, surge la pregunta por la diferencia entre los conceptos ELECTRÓN y TRAZO DE NIEBLA. El problema es, nuevamente, el de la individuación de conceptos. Si ambos conjuntos de representaciones perceptivas rastrean las mismas propiedades en el mundo, entonces ¿en qué sentido se puede sostener que son dos conceptos distintos? Prinz (2002) responde a esta pregunta apelando al comportamiento contrafáctico de uno y otro concepto. Mientras que TRAZO DE NIEBLA sirve para detectar trazos de niebla en todos los mundos

posibles, en mundos posibles en los cuales dispusiéramos de otros mecanismos o técnicas para detectar electrones (por ejemplo, si pudiéramos observar los electrones usando unos lentes especiales), ELECTRÓN detectaría algo diferente a trazos de niebla. Así, hay una diferencia en los contrafácticos que sustentan las relaciones nomológicas para ELECTRÓN y TRAZO DE NIEBLA. Las representaciones de trazos de niebla pueden servir como vehículos de ELECTRÓN en el mundo actual, pero no en otros mundos posibles en los que no se utilice la cámara de niebla para detectar electrones. Sin embargo, esta respuesta confirma que no hay manera de individuar ELECTRÓN y TRAZO DE NIEBLA en el mundo actual.

Sospecho que los científicos, sin embargo, no confunden uno y otro concepto y tampoco los usan indistintamente. De este modo, recurren a algo más que las representaciones perceptivas para individuarlos. Se me ocurren dos alternativas para explicar qué es lo que distingue a ambos conceptos y ambas resultan problemáticas para la teoría de proxitipos. En primer lugar, podría sostenerse que la diferencia entre uno y otro radica únicamente en las representaciones de las palabras del lenguaje natural correspondientes a uno y otro. De esta manera, se daría apoyo a la tesis de que las representaciones amodales o etiquetas no son prescindibles, sino que sirven como individuatoras. La estrategia de Prinz de incluir las palabras del lenguaje natural en la estructura de los proxitipos es una forma de incorporar, de manera encubierta, un sistema de símbolos etiquetadores para los conceptos.⁴⁶ Así, en el mundo actual, los detectores no son suficientes para la individuación de conceptos sin apelar a un sistema simbólico amodal como el lenguaje natural. La base perceptiva de ELECTRÓN y TRAZO DE NIEBLA es la misma, pero son conceptos distintos, de modo que debe haber algo más que los distinga, algo que no son las representaciones perceptivas. Así, la individuación a partir de palabras del lenguaje natural exhibiría la necesidad de recurrir a representaciones amodales.

Prinz sostiene que las representaciones de palabras del lenguaje natural son modales, dado que el lenguaje constituye un sistema de *input* dedicado: “La facultad del lenguaje procesa *inputs*, ocupa su propio circuito neuronal y utiliza su propio conjunto de reglas y representaciones” (Prinz 2002, p. 200). Ahora bien, considero que la cuestión de si las representaciones del lenguaje natural son modales o no plantea un dilema para la teoría de proxitipos. Por un lado, si estas representaciones fueran modales, entonces el argumento de parsimonia, que hemos visto en §1, fortalecido por las ventajas explicativas de los detectores,

⁴⁶ Destéfano (2009) presenta una crítica similar.

se vería debilitado. Una de esas ventajas es, según Prinz, la no arbitrariedad de los detectores en contraposición a los indicadores, que tienen una relación arbitraria con aquello que representan. Ahora bien, las palabras del lenguaje natural se relacionan arbitrariamente con las propiedades en el mundo que son sus referentes. De modo que incluir a las representaciones lingüísticas como representaciones modales que pueden funcionar como detectores implica perder las ventajas explicativas que el argumento de parsimonia fortalecido asignaba a los detectores. Por otro lado, si se admite que las palabras del lenguaje natural son representaciones amodales, y se las incorpora como parte constitutiva de los proxitipos, tal como hace Prinz al menos para algunos casos, esto equivale a admitir que los conceptos deben incorporar representaciones amodales. Ya sea que se considere a las representaciones del lenguaje natural como modales o amodales, su incorporación como parte constitutiva de los conceptos resulta ser una estrategia problemática para el neo-empirismo puesto que, como consecuencia, o bien se debilitan las ventajas explicativas de los detectores por sobre los indicadores, o bien se contradice la tesis de que todas las representaciones conceptuales son modales.

En segundo lugar, podría sostenerse que los científicos no confunden ELECTRÓN y TRAZO DE NIEBLA porque, independientemente de las diferentes palabras del lenguaje natural que se utilizan para expresarlos, se establece una relación causal entre los electrones y los trazos de niebla observables. Los trazos de niebla son el resultado de la interacción entre las partículas de vapor y las partículas introducidas en la cámara. Esto obliga a incorporar en el concepto ELECTRÓN un vínculo o enlace causal que lo conecte con TRAZO DE NIEBLA, pero, hemos visto en relación a las dificultades para individuar conceptos, que los enlaces en sí mismos no son representaciones perceptivas. Prinz ofrece una explicación de cómo podría representarse perceptivamente el concepto CAUSACIÓN a partir de representaciones de contigüidad y sucesión. Sin embargo, en el caso de ELECTRÓN, las relaciones de contigüidad y sucesión se dan entre entidades imperceptibles y sus consecuencias observables. De modo que aun si se introdujera con éxito una representación perceptiva de la relación causal entre electrones y trazos de niebla, ésta requeriría una representación de ELECTRÓN (puesto que se necesita representar la relación entre ELECTRÓN y TRAZO DE NIEBLA). Así, en este caso, la representación perceptiva ELECTRÓN presupone la representación de una relación causal y la representación de la relación causal, a su vez, presupone la representación de ELECTRÓN. En última instancia, es probable que los científicos que diseñaron los experimentos con cámaras de niebla ya pensarán en electrones antes de observar los trazos en dichas cámaras (Weinberg 2003). Esto

sugiere que tenían otras representaciones, no perceptivas, que les permitían pensar acerca de electrones.

Prinz explica el concepto abstracto de VERDAD apelando a una operación que relaciona o compara expectativas derivadas de nuestras creencias con las representaciones del mundo obtenidas en la experiencia. VERDAD se representaría simulando una relación de correspondencia entre representaciones. Así, se reduce el concepto VERDAD a una relación de confirmación de creencias por medio de representaciones perceptivas. Pero, como señala Sarnecki (2004), VERDAD no es equivalente a CONFIRMACIÓN. La relación de VERDAD podría explicarse como una relación de correspondencia entre creencias/oraciones y el mundo. Sin embargo, en la propuesta de Prinz lo que se estaría contrastando serían creencias/oraciones y representaciones perceptivas del mundo, de modo que VERDAD sería así una relación entre creencias y representaciones perceptivas, esto es, una relación meramente entre representaciones mentales. De esta manera, se dejaría afuera, precisamente, la concepción de VERDAD como una correspondencia con algo en el mundo.

Prinz (2004) responde a esta objeción que si bien VERDAD no se reduce a CONFIRMACIÓN, la operación de correspondencia sirve como una representación de VERDAD, puesto que esa operación es un detector confiable de VERDAD. Sin embargo, si bien puede servir para detectar casos de VERDAD, no representa por completo el concepto. El problema de la representación de VERDAD como CONFIRMACIÓN es, una vez más, el de la individuación de conceptos. Si uno no agota al otro, pero ambos se representan mediante la misma operación ¿cómo los distinguimos? Debe ser posible diferenciar, por ejemplo, los casos de ilusiones perceptivas. Imaginemos que estamos en la ruta bajo el sol y vemos adelante nuestro un charco de agua. Supongamos también que sabemos que no hay agua en la ruta, sino que se trata de una ilusión óptica. En este caso, somos capaces de pensar que no es verdad que hay agua en la ruta, aunque es verdad que tenemos una experiencia perceptiva visual de agua en la ruta. Para que esto sea posible, VERDAD debería estar representado como una relación con el mundo, y no sólo con nuestras experiencias perceptivas, a lo cual está limitada la operación de correspondencia.

Aquí el neo-empirismo se ve forzado, una vez más, a recurrir a las habilidades lingüísticas. Parte de estas habilidades consiste en comprender la relación entre las palabras y las cosas. Cuando adquirimos un lenguaje, “Aprendemos a etiquetar el mundo” (Prinz 2004, p. 422). No se trata de una comprensión teórica, sino de una capacidad. El concepto de VERDAD como una correspondencia con estados de cosas del mundo podría representarse a partir de esta habilidad lingüística de etiquetar el mundo. Podríamos representar una oración

verdadera como una etiqueta que indica un hecho en el mundo. Esta estrategia podría funcionar, pero juega a favor de la incorporación de representaciones amodales, dado que recurre explícitamente a representaciones que funcionan como etiquetas o indicadores.

Por otra parte, la estrategia que propone Prinz de representar VERDAD como CONFIRMACIÓN recurre a otros conceptos que difícilmente se puedan representar perceptivamente, tales como el de CORRESPONDENCIA o CORRECCIÓN, y se desplaza de esta manera el problema de la representación perceptiva de un concepto abstracto a la representación de otros conceptos abstractos. ¿Cómo sería una representación perceptiva de la operación de correspondencia? ¿Cómo se representa perceptivamente la correspondencia entre dos representaciones perceptivas distintas, o entre una creencia y una experiencia? Sin una respuesta a esto que pueda prescindir de las representaciones simbólicas amodales, el neo-empirismo no logra satisfacer el requisito de alcance.

Los ejemplos de aplicación de las estrategias sugeridas por Prinz muestran que la representación de conceptos teóricos y abstractos a partir de representaciones perceptivas sigue siendo un desafío para el neo-empirismo, puesto que para ampliar el alcance de la teoría, los proxitipos deben incorporar como elementos constitutivos de los conceptos algo que no es en sí mismo perceptivo, ya sea enlaces entre representaciones o indicadores. En última instancia, los problemas relacionados con el alcance dependen de los problemas señalados para la individuación del contenido. Como vimos para los ejemplos de ELECTRÓN y VERDAD, la teoría no puede diferenciar ELECTRÓN de TRAZO DE NIEBLA, ni VERDAD de CORRECCIÓN, sin recurrir a un nivel distinto del de las representaciones perceptivas. Esto se debe a las dificultades para determinar el contenido de esos conceptos sin apelar a indicadores, como vimos en relación al requisito de contenido.

3.3 Composicionalidad

El neo-empirismo también tiene ciertas dificultades para dar cuenta de la composicionalidad. Con respecto a este requisito, sin embargo, la teoría de proxitipos presenta, a la vez, ventajas y desventajas. Por un lado, Prinz ofrece un modelo de combinación de conceptos que sugiere cómo intervienen los distintos tipos de estructuras conceptuales para formar nuevas combinaciones de conceptos y que permite explicar los casos de propiedades emergentes y pérdida de propiedades de los conceptos iniciales. Pero, por otro lado, este modelo no consigue abarcar todos los casos de combinación, como sostendré a continuación.

Prinz (2002) elabora un modelo de combinación de conceptos en tres etapas, el modelo de *Recuperación Composición y Análisis* (RCA), que incorpora distintas estrategias de combinación, entre las cuales se encuentra la composicionalidad. Recordemos que Prinz entiende el requisito de composicionalidad en sentido débil, i.e. sostiene que, dado que tanto la productividad como la sistematicidad son capacidades, y la composicionalidad es la mejor explicación para ellas, esto implica solamente que se debe tener la capacidad de combinar conceptos de modo composicional, pero no que éste es el único modo en que se combinan los conceptos (Prinz 2002). Así, sólo debería mostrar que es posible combinar proxitipos de modo composicional. Pues bien, para mostrar que este tipo de combinación es posible, Prinz defiende la composicionalidad de los prototipos. Los proxitipos no son idénticos a prototipos, pero pueden tener, entre otras, una estructura prototípica. En especial, los proxitipos por *default*, i.e. los que se activan en ausencia de un contexto específico, son estructuras estables de tipo prototípico. De manera que el peso de la defensa recae sobre la posibilidad de combinar prototipos que, como vimos en el capítulo 2, resulta problemática.

Como sostendré, la defensa indirecta de la composicionalidad de los proxitipos a través de la composicionalidad de los prototipos no resulta suficiente. Por un lado, esta defensa no consigue hacer frente a las críticas que se han esgrimido contra la composicionalidad de los prototipos y, por otro lado, el modelo no resulta suficiente para mostrar la composicionalidad de los proxitipos en general. Para que la defensa indirecta funcione, Prinz debería mostrar o bien que todo proxitipo tiene una parte prototípica o bien que existe un modelo de composición para los proxitipos que no exhiben estructura prototípica, pero no muestra ninguna de estas dos cosas.

El RCA tiene tres etapas. La primera es la de *recuperación*, en la cual se busca en la memoria información relevante para representar un concepto compuesto. En esta etapa se puede obtener una representación del concepto compuesto de dos formas distintas, ninguna de las cuales apela a la composicionalidad. Por un lado, podrían reactivarse representaciones de ejemplares del concepto compuesto. Esta sería una manera no composicional de arribar a ese concepto, en tanto no se parte de conceptos más simples, sino que directamente se activa un ejemplar. Por otro lado, podría realizarse un proceso de cruzamiento de listas de rasgos para ver si se encuentra una representación de un ejemplar que caiga bajo los dos conceptos que se quiere combinar. Por ejemplo, si se quiere representar CUCHARA DE MADERA puede buscarse un ejemplar que tenga los rasgos correspondientes a CUCHARA y el rasgo DE MADERA. De esta manera se explica la aparición de rasgos emergentes: estos rasgos provienen de las representaciones particulares de ejemplares en la memoria.

Si esta etapa falla, i.e. si no se encuentran ejemplares del concepto compuesto en la memoria, se pasa a una segunda etapa, la de *composición*. Esta etapa permite combinar conceptos siguiendo reglas de combinación, algunas de las cuales permitirían satisfacer el requisito de composicionalidad. Prinz presenta una heurística para combinar conceptos con estructura prototípica basada en los procesos de alineación e integración, en la que se puede reconocer la influencia del *Modelo de Modificación Selectiva* de Smith *et al.* (1988), reconstruido brevemente en el capítulo 2. En el modelo que Prinz propone, la alineación consiste en poner en paralelo los rasgos comunes (o similares) a los dos conceptos, mientras que la integración consiste en ajustar los valores de esos rasgos comunes a ambos. Esta etapa podría ser, según Prinz, puramente composicional (sin recurrir a representaciones de ejemplares en la memoria ni a conocimiento de trasfondo), cuestión a la que volveré más adelante.

Por último, hay una etapa de *análisis* en la cual se evalúa si el resultado de la combinación es coherente. Para esto entran en juego el conocimiento de trasfondo y los componentes estructurados como teorías que pueden incluir los proxitipos, “se deben salvar las brechas, se deben extraer inferencias, y se deben eliminar o explicar las inconsistencias aparentes” (Prinz 2002, p. 306). Para ilustrar lo que ocurre en esta etapa, Prinz toma un ejemplo de Murphy que resulta muy esclarecedor: SIERRA ANTIGUA tiene la propiedad emergente OXIDADA. Murphy explica esta propiedad como el resultado de una inferencia estadística: “las sierras son típicamente de metal; los objetos antiguos son viejos; y el metal viejo típicamente está oxidado; de allí que una sierra antigua deba estar oxidada” (Prinz 2002, p. 307). Así, el proceso de análisis también permite explicar las propiedades emergentes, que pueden ser resultado, por ejemplo, de inferencias a partir de conocimiento de trasfondo.

Considero que este modelo no resulta suficiente para mostrar que los proxitipos se pueden combinar de modo composicional. La defensa que Prinz ofrece de la composicionalidad de los proxitipos consiste en una defensa indirecta: se ocupa de responder a las objeciones respecto de la composicionalidad de los prototipos. ¿Cómo justifica esta estrategia? Prinz introduce la idea de los proxitipos por *default* para responder al requisito de publicidad. Si pretendemos que los conceptos tengan un papel en la explicación del comportamiento y la comunicación, entonces deben ser compartibles. Para poder predecir, por ejemplo, similitudes en el comportamiento en relación a los perros, es necesario que compartamos el concepto PERRO. Respecto de la comunicación, ésta funciona porque asociamos los mismos conceptos con las palabras, así nos entendemos porque cuando

usamos, por ejemplo, la palabra “perro”, la usamos como asociada al mismo concepto PERRO (Prinz 2002, p. 14).

El problema que aparentemente tendrían los proxitipos para satisfacer el requisito de publicidad es que son sensibles al contexto. La manera en que se representa EDIFICIO, por ejemplo, será diferente en Dublin, donde no suelen superar los cinco pisos, y en Dubai, donde tienen un promedio de 28 pisos de altura. Los proxitipos por *default* son las representaciones de una categoría que se activarían si no se diera ningún contexto particular. Los rasgos contenidos por los proxitipos por *default* dependen de la frecuencia con la que representamos una categoría como teniendo ese rasgo. Así, el contenido de estos proxitipos está estructurado en rasgos salientes, típicos y diagnósticos ponderados, al igual que las estructuras de los prototipos y, por lo tanto, se tornan blanco de las críticas a la composicionalidad de los prototipos, de las que me ocupé en el capítulo 2: los casos de conceptos sin prototipo, como CIUDADES AMERICANAS SITUADAS EN LA COSTA ESTE UN POCO AL SUR DE TENNESSE, y el hecho de que el contenido de los conceptos compuestos no parece ser una función de los contenidos de los prototipos de los conceptos iniciales, evidenciado en algunos casos por propiedades emergentes que no se encontraban en los conceptos iniciales y en otros por la pérdida de propiedades de los conceptos iniciales.

Con respecto a la primera objeción, Prinz afirma que tener un prototipo es tener una estructura “en grados de tipicidad”, i.e. debe haber instancias que son más típicas de ese concepto que otras. De esta manera, para mostrar que un concepto tiene un prototipo, sólo es necesario mostrar que hay instancias más típicas de ese concepto que otras. Prinz propone que, por ejemplo, para el concepto ABUELAS LA MAYORÍA DE CUYOS NIETOS SON DENTISTAS, una mujer que tiene 8 nietos, de los cuales 6 son dentistas será un ejemplar más típico de este concepto que una mujer que tiene 5 nietos, de los cuales 3 son dentistas. Y la misma respuesta valdría para NO GATO o para ROSA SI ES UN CUADRADO. Un puma sería un ejemplar mucho menos típico de NO GATO que una mesa, y un cuadrado rosa sería un ejemplar mucho más típico de ROSA SI ES UN CUADRADO que un cilindro verde o una pelota violeta.

Esta respuesta apela a intuiciones. Sin embargo, la pregunta por si este tipo de conceptos presenta una estructura prototípica o “en grados” debería zanjarse empíricamente. Prinz reconoce esto, pero considera que un breve ejercicio de introspección resulta suficiente para dar un indicio de que estos conceptos presentan una estructura en grados. Aún si decidiéramos conceder la intuición a la que Prinz apela, es cuestionable el tipo de gradación que ofrece para conceptos como los vistos. ¿Qué resulta más típico de ABUELAS LA

MAYORÍA DE CUYOS NIETOS SON DENTISTAS? Prinz parece considerar que toda la tipicidad de este concepto pasa por una estructura en grados de MAYORÍA (6/8 resulta una instancia de mayoría más prototípica que 3/5, aunque sólo es una mayoría “mayor” y no por eso necesariamente más prototípica). De manera que no queda claro que conceptos como NO GATO o ABUELAS LA MAYORÍA DE CUYOS NIETOS SON DENTISTAS puedan considerarse como prototipos y menos aún como el resultado de componer otros prototipos.

Con respecto a la segunda objeción, el modelo RCA consigue explicar los casos de propiedades emergentes y los casos de pérdida de propiedades de los conceptos iniciales apelando a representaciones distintas de los prototipos, en particular a ejemplares y a estructuras teóricas. En este sentido, ofrece una respuesta a esta objeción. Sin embargo, esta respuesta elude la cuestión de si los prototipos se pueden combinar composicionalmente, i.e. si se puede determinar el contenido de un concepto compuesto únicamente a partir de los prototipos constituyentes y determinadas reglas de combinación y, a la inversa, si, frente a un concepto compuesto, se puede identificar la contribución de cada uno de los prototipos constituyentes. Como vimos en el capítulo 2, si bien los modelos de combinación de prototipos consiguen explicar algunos casos sencillos apelando a reglas de combinación, su alcance resulta limitado, y no pueden extenderse a casos más complejos. El modelo RCA no es una excepción y las reglas que propone no se alejan de las del modelo de modificación selectiva.

Por otra parte, Prinz enfatiza que los proxitipos son construcciones altamente variables en la memoria de trabajo y que lo que se activa puede ser tanto una estructura prototípica como un modelo mental o una imagen de una palabra del lenguaje natural. Señala que: “Un proxitipo puede ser una representación multimodal detallada, un modelo visual simple, o incluso una representación de una palabra (e.g. una imagen auditiva de la palabra “perro”). El contexto determina qué proxitipo se usa en la memoria de trabajo en cualquier ocasión dada” (Prinz 2002, p. 149). Esto difícilmente sea compatible con la afirmación de que estas estructuras se pueden combinar de manera sistemática. La sistematicidad del pensamiento es la capacidad de tener ciertos pensamientos *si* se tienen otros o, dicho de otra manera, el hecho de que no es posible ser capaz de pensar aRb y no ser capaz de pensar bRa . Como vimos en el capítulo 1, Fodor y Pylyshyn (1995) señalan que la mejor manera de explicar este fenómeno es apelando a la estructura del pensamiento. Pero esto no es todo. También hay que suponer que la contribución que hace a en aRb y en bRa es la misma. Sin embargo, en el caso del neo-empirismo, se trata de representaciones sensibles al contexto, y precisamente esta variabilidad consiste en que hacen contribuciones distintas en contextos

distintos. Esto trae nuevamente el problema de la individuación de conceptos. Si se pudiera mostrar que dos proxitipos distintos son iguales en tipo, se podría sostener que en ese sentido contribuyen con lo mismo a la representación compleja. Pero, como sostuve más arriba, esto no es posible. De modo que no se puede explicar la sistematicidad a partir de estructuras como los proxitipos.

Supongamos por mor del argumento de Prinz que cuando se activa una palabra o una estructura de tipo prototípico en la memoria de trabajo, estas representaciones son combinables de manera productiva y sistemática. Todavía restaría mostrar que el resto de las estructuras que puede presentar un proxitipo se puede combinar de modo composicional. Esto no resulta plausible, por lo menos, para los conjuntos de ejemplares. Así, por un lado, dado que la respuesta a los casos de conceptos sin prototipo resulta débil, es lícito afirmar que Prinz no logra mostrar que todos los proxitipos tienen una parte prototípica. Y, por otra parte, tampoco explica cómo podrían combinarse de modo composicional los conjuntos de ejemplares o las estructuras teóricas. Así, los proxitipos no parecen poder satisfacer el requisito de composicionalidad.

Por último, la combinación de representaciones perceptivas, en tanto simulación de estados perceptivos, no parece capaz de dar cuenta del pensamiento proposicional. Por ejemplo, tener un pensamiento proposicional como que los aviones tienen alas, equivaldría a reactivar representaciones perceptivas de un avión y vincularlas con una simulación de alas. Sin embargo, no queda claro cómo esta simulación recoge la idea de que los aviones *tienen* alas, en lugar de que los aviones *son* alas, y resulta dudoso que esta diferencia se pueda establecer sin apelar a un sistema de representaciones lingüístico o simbólico (Machery 2006).

Una individuación de ese tipo resulta necesaria para poder dar cuenta del pensamiento proposicional y de los procesos de razonamiento formal. El problema es que las simulaciones no permiten distinguir distintos tipos de predicados, ya que no es distinguible o “separable” la contribución que hace cada proxitipo a la simulación completa. Para este tipo de individuación sería necesario que las representaciones perceptivas tuvieran una estructura sintáctica, que no tienen. Como vimos en el capítulo 4, la relación de predicación se representa adecuadamente por medio de representaciones discursivas, las cuales permiten distinguir constituyentes sintácticos canónicos (Fodor 2008). Las representaciones discursivas tienen una forma lógica que hace explícita la estructura de constituyentes, puesto que explicita cómo contribuye cada parte a la interpretación del símbolo entero. Las representaciones perceptivas, en cambio, sólo admiten una distinción de partes, i.e. cada parte

contribuye de la misma manera a la interpretación, de modo que no tienen una estructura de constituyentes canónicos y, por lo tanto, no pueden representar las relaciones predicativas.

Prinz intenta explicar la predicación entre representaciones perceptivas por medio de un enlace predicativo con redes de memoria. Supongamos que se activa una simulación que consiste en una representación perceptiva de un insecto de ocho patas caminando por la pared. Esta será una simulación de una araña si está predicativamente ligada a una red de memoria a largo plazo de ARAÑA, es decir, si hay una disposición a transferir a esa representación perceptiva particular las propiedades o rasgos almacenados en la red de memoria de ARAÑA. Ahora bien, en este caso, el enlace predicativo es equivalente a la relación de pertenencia a una categoría. En el ejemplo de los aviones, en cambio, que los aviones tienen alas no puede implicar un vínculo de predicación equivalente a la pertenencia a una categoría, ya que no se está representando que los aviones pertenecen a la categoría ALAS (ni que las alas pertenecen a la categoría AVIONES). Y se quiere, precisamente, poder distinguir el pensamiento de que los aviones *tienen* alas del pensamiento de que los aviones *son* alas.

De modo que para relacionar AVIÓN y ALAS, hay que recurrir a otro tipo de enlace. Podría ser al enlace de unión, que permite asociar representaciones perceptivas de distintas modalidades. Sin embargo, esta alternativa supone que lo que relaciona o une a las distintas representaciones perceptivas en una simulación son enlaces del mismo tipo que los que asocian las distintas representaciones perceptivas en una misma red en la memoria a largo plazo. Pero esto no es algo que se especifique en la teoría de proxitipos, sino que Prinz simplemente sostiene que en una simulación se reactivan de manera conjunta distintas representaciones perceptivas que estaban almacenadas en distintas redes de la memoria a largo plazo. Por otra parte, recurrir a distintos enlaces entre representaciones perceptivas para explicar cómo se pueden representar de distinto modo SER y TENER, supondría que no hay un proxitipo para SER ni TENER, lo cual significaría que no hay un concepto SER ni un concepto TENER.

Así, para poder distinguir entre el pensamiento de que los aviones tienen alas y el pensamiento de que las alas tienen aviones no es suficiente con la simple asociación de simulaciones de representaciones perceptivas (Machery 2006). Esto requiere representaciones que establezcan distinciones más finas en la relación de predicación, tales como las representaciones discursivas. En consecuencia, la individuación de pensamientos proposicionales parece requerir representaciones simbólicas o lingüísticas.

El análisis del neo-empirismo sugiere que la postulación de representaciones perceptivas como los únicos vehículos conceptuales no resulta suficiente para dar cuenta de los requisitos fundamentales de una teoría de conceptos, por lo que sugeriré que resultará ventajoso abandonar el principio de la uniformidad no sólo en relación a las estructuras semánticas, sino también en relación a los vehículos conceptuales. En el próximo capítulo reconstruiré las tesis principales de las teorías híbridas y pluralistas, así como su oposición al eliminativismo. Esto me permitirá delinear las características generales de la propuesta pluralista que desarrollaré en el capítulo 7.

Capítulo 6

Las teorías híbridas, pluralistas y eliminativistas

En los capítulos previos analicé las ventajas y los límites explicativos de los enfoques que adhieren al principio de uniformidad, i.e. que atribuyen una constitución uniforme a todos los conceptos, ya sea respecto de su estructura interna, como es el caso de los enfoques de definiciones, prototipos, ejemplares, teorías y el atomismo, o en relación a su vehículo, como es el caso de la teoría del atomismo conceptual y la teoría de proxitipos. Este análisis indica que ninguno de ellos está en condiciones de acomodar los requisitos explicativos fundamentales, pero que, sin embargo, cada uno resulta adecuado para dar cuenta de algunos de ellos. En el marco de los debates entre estos enfoques, se generó un cuerpo de evidencia empírica, así como una serie de argumentos que tuvieron como efecto el debilitamiento de las distintas teorías que adoptan el principio de la uniformidad. Como consecuencia, muchos autores propusieron combinaciones de algunas de las estructuras, como teorías y prototipos (Keil 1989, Keil *et al.* 1998) o definiciones y prototipos (Osherson & Smith 1981), dando lugar a las primeras teorías híbridas. Un ejemplo de este tipo es la teoría de proxitipos, analizada en el capítulo anterior, la cual abandona el principio de uniformidad en relación a la estructura conceptual interna, aunque lo conserva respecto de los vehículos conceptuales.

Actualmente, predominan los intentos por unificar distintos tipos de estructuras conceptuales en una misma teoría de conceptos, y el debate principal se da entre enfoques que defienden distintos modos de articular las diversas estructuras. Usualmente, se agrupa las teorías que abandonan el principio de uniformidad bajo uno de dos enfoques: el híbrido y el pluralista. En este capítulo sostendré que, si bien las distintas teorías que se agrupan en estos enfoques son diferentes en sus detalles y modelos particulares, los enfoques en sí no están en competencia, sino que comparten sus tesis centrales. Ofreceré un diagnóstico tentativo acerca de por qué se adopta la distinción entre teorías híbridas y pluralistas. Asimismo, analizaré el debate entre estas posturas y el eliminativismo, según el cual la clase “concepto” constituye una clase heterogénea y debe ser eliminada de la ciencia cognitiva. Así, los objetivos de este capítulo son, primero, elucidar las tesis centrales de las teorías llamadas híbridas y pluralistas y mostrar que hay acuerdo entre ellas, de modo de explicitar las tesis que adoptaré como punto de partida para el desarrollo del pluralismo amplio en los próximos capítulos. Segundo, mostrar que este tipo de propuesta no conduce al eliminativismo de la noción de “concepto”. Ofreceré posibles respuestas a los argumentos eliminativistas, que no implican un retorno al principio de uniformidad.

1. Propuestas híbridas y pluralistas

Resulta usual la reconstrucción de los debates actuales entre enfoques que postulan una multiplicidad de representaciones conceptuales heterogéneas distinguiendo dos enfoques predominantes, uno híbrido y otro pluralista (Machery, 2009; 2010; Machery & Seppälä, 2011; Weiskopf, 2009a). Tanto las propuestas híbridas como las pluralistas reconocen que en los procesos cognitivos superiores intervienen estructuras conceptuales heterogéneas y que no hay un único tipo de estructura que pueda dar cuenta de todos los fenómenos que las teorías de conceptos deberían explicar. En general, las distintas propuestas incorporan estructuras prototípicas, teóricas y conjuntos de ejemplares, puesto que son las estructuras para las que se dispone de mayor evidencia, aunque generalmente consideran la posibilidad de que haya otras estructuras, por ejemplo, ideales (Machery, 2009; Weiskopf, 2009a). Usualmente, se asume que la diferencia entre las versiones híbridas y las pluralistas está dada por la relación que postulan entre esas estructuras.

Las teorías híbridas proponen que los conceptos tienen partes, constituidas por estructuras diferentes (Anderson & Betz 2001, Keil *et al.* 1998, Margolis & Laurence 2007, 2010, Osherson & Smith 1981). Así, por ejemplo, Margolis & Laurence (2003, 2007, 2010) sostienen que los conceptos tienen una estructura semántica que determina la referencia y una “estructura para el procesamiento” que puede contener prototipos, ejemplares y teorías, y es la que explica cómo intervienen los conceptos en diversos procesos mentales. Keil (1989, Keil *et al.* 1998), por su parte, sostiene que los conceptos tienen una parte prototípica y otra teórica. En cambio, las teorías pluralistas proponen que cada categoría está representada por diversos conceptos independientes, cada uno de los cuales posee una estructura diferente. Así, habría prototipos, ejemplares y teorías, cada uno representando de modo independiente una categoría (Machery 2009, Weiskopf 2009a).⁴⁷

De este modo, usualmente se interpreta los enfoques híbrido y pluralista como dos enfoques en competencia y, más allá de las diferencias particulares entre ellas, se agrupa las distintas teorías que abandonan el principio de uniformidad en alguno de esos dos enfoques. A continuación intentaré mostrar que estos enfoques no están en competencia respecto de sus compromisos centrales, sino que las distintas teorías contemporáneas que postulan una multiplicidad de estructuras conceptuales heterogéneas parten de ciertas tesis en común y se diferencian por los tipos de estructuras que postulan (algunas postulan prototipos, ejemplares

⁴⁷ Estas teorías no se comprometen con que éste sea el caso para todas las categorías, podría haber algunas que sólo representemos por medio de, e.g., teorías (como los electrones).

y teorías, mientras que otras postulan, además, representaciones lingüísticas y símbolos atómicos) y los modelos específicos que proponen. Sostendré, por un lado, que los argumentos principales contra las teorías híbridas les atribuyen compromisos que estas teorías no adoptan y, por otro lado, que ambos enfoques comparten ciertas tesis.

Discutiré fundamentalmente los criterios que ofrece Machery para distinguir entre teorías híbridas y pluralistas. Tal como argumenta Machery (2009; 2010), si no se explicitan condiciones necesarias y suficientes para que una estructura sea parte de un concepto (en lugar de constituir un concepto independiente), no es posible establecer una diferencia significativa entre las propuestas híbridas y las pluralistas. Sin embargo, las condiciones que usualmente se atribuyen a las teorías híbridas no parecen ser tesis que estas teorías adopten. Machery (2009, Machery & Seppälä 2011) formula dos condiciones que asume implícitas en las teorías híbridas, la condición de coordinación y la de asociación. En §2 presentaré estas dos condiciones y argumentaré que las propuestas que actualmente se denominan híbridas no se comprometen con ellas, por lo menos no en el modo en que las interpreta Machery. En §3 discutiré otro argumento contra las teorías híbridas, presentado por Weiskopf (2009a) y sostendré que, al igual que los argumentos de Machery, se basa en una interpretación no adecuada de las teorías híbridas que denominaré “arquitectural”. Concluiré que las condiciones de coordinación y asociación no ofrecen una distinción adecuada entre las teorías híbridas y pluralistas, puesto que, si bien fueron propuestas para distinguirlas, terminan acercándolas. Aparentemente, tanto el enfoque híbrido como el pluralista coinciden en rechazar la condición de coordinación y en aceptar una versión semántica de la condición de asociación. De este modo, sostendré que la oposición entre teorías híbridas y pluralistas es, con respecto a sus tesis generales, sólo aparente.

En §4 presentaré un diagnóstico tentativo acerca de los motivos por los cuales ciertos autores, especialmente los defensores de las llamadas teorías híbridas, se resisten a reconocer que el debate es en última instancia terminológico. Sostendré que se trata de un intento por resistir el eliminativismo de Machery. Me ocuparé de reconstruir la estrategia eliminativista de Machery y ofreceré un modo de rechazarla que no implica conceder la distinción entre teorías híbridas y pluralistas, pero tampoco negar que los conceptos constituyen una clase heterogénea.

2. Las condiciones de coordinación y asociación

Como argumenta Machery (2009, 2010), para poder diferenciar entre propuestas híbridas y pluralistas, si es que hay diferencias entre ellas, es necesario determinar cuáles son

las condiciones necesarias y suficientes para que una estructura sea parte de un concepto o un concepto independiente en sí mismo. Sin tales condiciones, no habría una oposición genuina entre los enfoques híbrido y pluralista. Si, desde los enfoques híbridos, se pretende que las distintas estructuras formen parte de un mismo concepto, entonces esas partes deben comportarse de algún modo diferente a los conceptos independientes. Si no se postulara ninguna marca específica para “ser parte de un concepto”, entonces la distinción entre enfoques híbridos y pluralistas pasaría a ser meramente terminológica (Machery, 2009; 2010; Machery y Seppälä, 2011).

Las teorías híbridas no han propuesto formalmente condiciones que las diferencien de las pluralistas y no siempre está claro en qué sentido consideran que las distintas estructuras son partes de un mismo concepto. Según Machery (2009), estas teorías presuponen dos condiciones para que una estructura sea parte de un concepto, la *coordinación* y la *asociación*. La primera establece que las distintas partes de un mismo concepto no pueden producir resultados inconsistentes (por ejemplo, no pueden producir juicios de categorización inconsistentes). La segunda establece que las distintas partes de un concepto están conectadas: cuando una parte está en uso, las otras partes “están preparadas” para ser usadas también. Sin embargo, resulta cuestionable que éstas sean las condiciones que tienen en mente los defensores de las teorías híbridas.

La distinción que propone Machery entre teorías híbridas y pluralistas no es la de un observador neutral, que pretende reconstruir un debate del que no participa, sino que forma parte de su defensa del eliminativismo de conceptos, que analizaré en §3. Quisiera adelantar sin embargo parte de la estrategia general de Machery. Ésta consiste, en un primer paso, en distinguir entre teorías híbridas y pluralistas, a partir del supuesto compromiso de las teorías híbridas con las condiciones de coordinación y asociación. En un segundo paso, Machery presenta resultados empíricos que irían en contra de la condición de coordinación y concluye que hay razones para preferir las teorías pluralistas en lugar de las híbridas. Machery afirma que si las teorías híbridas no adoptan las condiciones de coordinación y asociación, entonces no son distintas de las pluralistas. Y, si las aceptan, entonces hay evidencia significativa en su contra. Por lo tanto, las teorías pluralistas se deben preferir por sobre las híbridas. En un tercer paso, del que me ocuparé en detalle en §3, Machery argumenta que la tesis pluralista, según la cual la clase de los conceptos incluye una multiplicidad de representaciones heterogéneas conduce al eliminativismo de la noción de “concepto”.

En las próximas secciones presentaré una serie de objeciones a esta estrategia. En §2.1 discutiré la condición de coordinación. Señalaré, por un lado, que las teorías híbridas no

parecen aceptar esta condición y, por otro lado, que la evidencia contra esta condición tampoco resulta determinante. En §2.2 me ocuparé de la condición de asociación. Sostendré que, mientras que Machery parece interpretar esta condición en un sentido que denomino arquitectural, ésta debería ser entendida en un sentido semántico. Interpretada de este modo, la condición de asociación está presupuesta en las principales teorías híbridas. Sin embargo, sostendré que algunas versiones pluralistas parecen presuponer una condición similar.

2.1 La condición de coordinación

Según la condición de coordinación, las partes de un concepto están coordinadas: “las partes de un concepto dado no producen resultados inconsistentes, por ejemplo, juicios de categorización inconsistentes” (Machery 2009, p. 64). Así, según esta condición, si los conceptos tuvieran una estructura híbrida, no deberíamos asentir a juicios inconsistentes como “El café es agua” y “El café no es agua”. Machery organiza su discusión de las teorías híbridas en torno a las consecuencias que se siguen de la adopción de esta condición y presenta evidencia empírica contra estas consecuencias. Sin embargo, tanto la condición como la evidencia han resultado muy cuestionadas (Gonnerman & Weinberg, 2010; Keil, 2010; Margolis & Laurence, 2010; Scarantino, 2010). En primer lugar, defensores de teorías híbridas, como Margolis y Laurence (2010) o Keil (2010), rechazan la condición de coordinación como una condición necesaria para que diversas estructuras sean partes de un concepto, por ser una condición demasiado fuerte que conduciría a consecuencias que estas teorías pretenden evitar. En segundo lugar, aún si las teorías híbridas aceptaran esta condición, la evidencia contra la coordinación es muy cuestionable. Desarrollaré estas dos objeciones en orden.

La principal crítica a la condición de coordinación señala que ésta es demasiado fuerte, dado que obligaría a tratar como conceptos independientes a tipos de estructuras que los psicólogos usualmente tratan como unificadas. Keil (2010) señala que, por ejemplo, el uso de ALTO en tareas de categorización puede conducir a juicios aparentemente inconsistentes en función del contexto en que se evalúa si un individuo es alto o no. Sin embargo, postular que hay un concepto independiente por cada uno de los “micro-contextos” que pueden conducir a tales juicios aparentemente contradictorios llevaría a una proliferación extrema de conceptos. Keil sostiene que estos casos se suelen explicar como aplicaciones de un mismo concepto y no como usos de conceptos distintos. Por otra parte, Gonnerman y Weinberg (2010) señalan que la condición de coordinación implicaría que los conjuntos de ejemplares no puedan ser tratados como un todo conceptual. Ciertos estudios sugieren que cambios

contextuales en la perspectiva del sujeto (por ejemplo, desde la perspectiva de un artista a la de un biólogo) pueden dar lugar a distintos resultados al categorizar un ítem debido a que la adopción de una perspectiva diferente lleva a activar distintos ejemplares para comparar en la categorización (Braisby 2005). Determinados cambios contextuales pueden inducir la activación de distintos ejemplares de un conjunto que representa una categoría, y estos cambios dan lugar a respuestas diversas frente a una tarea de categorización.

Machery (2010) reconoce la fuerza de la objeción de Gonnerman y Weinberg, pero sólo ofrece como respuesta que en última instancia esto deberá probarse empíricamente. Si efectivamente se diera este efecto para los conjuntos de ejemplares, estos ya no podrían tratarse como un todo conceptual, sino que deberían tratarse como conceptos independientes. Pero en ese caso, resultaría difícil integrar los ejemplares en las teorías pluralistas de conceptos, dado que las propuestas disponibles de procesamiento de ejemplares los tratan como un todo (Medin & Schaffer 1978, Nosofsky 1986, Nosofsky & Palmeri 1997). De modo que la condición de coordinación podría incluso jugar en contra del tipo de teoría pluralista que Machery propone defender, en tanto el pluralismo de Machery estaría incorporando como “ejemplares” algo distinto de lo que las propuestas actuales de ejemplares defienden.

Por otra parte, suponiendo que la tesis de coordinación efectivamente resultara central para las teorías híbridas, la evidencia en su contra está lejos de ser concluyente. Machery y Seppälä (2011) sostienen que si disponemos de distintos conceptos correferenciales para, e.g. tomate, es probable que “tomate” sea polisémica, i.e. que esté asociada a distintos significados, correspondientes a cada uno de los conceptos TOMATE_P, TOMATE_E y TOMATE_T.⁴⁸ Esta es la hipótesis de la polisemia. El estudio que proponen pretende dar apoyo a esta hipótesis y mostrar que la polisemia permite que aceptemos juicios inconsistentes, lo cual no debería ocurrir si las distintas estructuras que representan una categoría cumplieran con la condición de coordinación. Así, el estudio pretende brindar apoyo a las teorías pluralistas contra las híbridas. El estudio consiste en pedirle a un grupo de individuos que evalúen en una escala del 1 al 7 en qué medida acuerdan con determinados juicios de clasificación. Los juicios se presentan de a pares, y siempre incluyen una afirmación seguida de su negación, por ejemplo “En un sentido, los tomates son verduras” y “En un sentido, los tomates no son verduras”. Se presentan 6 pares blanco y 3 de control. Al

⁴⁸ Utilizaré los subíndices P para prototipo, E para conjunto de ejemplares y T para teoría. Así, TOMATE_P refiere al prototipo de *tomate*, indistintamente de si se lo considera una parte del concepto TOMATE o un concepto independiente.

confeccionar los pares de juicios blanco, los autores seleccionaron o bien predicados cuyas extensiones son similares a los prototipos del otro predicado, pero que no caen bajo la extensión de la teoría que también asumen asociada a ese predicado (e.g. “tomates” y “verduras”), o bien predicados cuyas extensiones no son similares a los prototipos del otro predicado, pero caen bajo su extensión en virtud de la teoría asociada (e.g. “pingüinos” y “aves”). Para los pares de control, en cambio, seleccionaron predicados cuyas extensiones caen bajo el otro predicado sin importar si se tiene en cuenta el prototipo o la teoría asociados (e.g. “En un sentido los leones son animales”). Los resultados del estudio indican un amplio porcentaje de acuerdo con los pares de juicios blanco que son, según Machery y Seppälä, aparentemente inconsistentes.

Los autores interpretan estos resultados como indicios de que las distintas estructuras semánticas correferenciales, que según las teorías híbridas serían partes de un mismo concepto, no cumplen con la condición de coordinación, puesto que permiten juicios aparentemente inconsistentes. Así, estos resultados servirían como evidencia para preferir las teorías pluralistas por sobre las híbridas. Machery y Seppälä reconocen que estos resultados sólo ofrecen un apoyo tentativo a favor de las teorías pluralistas y anticipan una serie de objeciones. Como argumentaré a continuación, las respuestas a las objeciones, en lugar de fortalecer el argumento contra las teorías híbridas, lo debilitan aun más. Una de las objeciones posibles está centrada en la presencia de la expresión “en un sentido” en los pares de afirmaciones que los participantes debían evaluar. Señalan que se podría objetar que esta aclaración invita a evaluar la adecuación pragmática de las afirmaciones, en lugar de evaluar su valor de verdad. A esta objeción responden, primero, que los comentarios y aclaraciones que aportaron los participantes en el estudio indican que están evaluando el valor de verdad. Segundo, señalan que en los ejemplos de control (e.g. “En un sentido, los leones son animales” y “En un sentido, los leones no son animales”) la presencia de la expresión “en un sentido” no motivó el asentimiento a juicios aparentemente contradictorios. Sin embargo, considero que estas respuestas no resultan suficientes para superar la objeción.

En relación a la primera respuesta, no resulta evidente a partir de los comentarios de los individuos que estos estuvieran evaluando el valor de verdad de las afirmaciones y no aspectos más bien pragmáticos. Por ejemplo, algunos de los comentarios que dejaron en relación al par “En un sentido, los tomates son verduras” y “En un sentido, los tomates no son verduras” fueron: “Tienen semillas, entonces son frutas –pero el uso que hacemos de ellos al cocinar es más similar al modo en que cocinamos otras verduras”, “Creo que son frutas, pero los comemos como verduras, entonces, en un sentido, los tomates son verduras” o

“Botánicamente son frutas, pero no son dulces y se usan fundamentalmente en platos salados y en conjunción con otras verduras” (Machery & Seppälä 2011, p. 119). Estos comentarios sugieren, contra la lectura de Machery y Seppälä, que los individuos están considerando más bien que los tomates no son verduras, pero que, puesto que los usamos como tales, hay un sentido pragmático en el que resulta adecuado afirmar que los tomates son verduras.

En relación a la segunda respuesta, considero que las evaluaciones de los ejemplos de control no permiten inferir si los individuos están llevando a cabo una evaluación pragmática o de valor de verdad. Machery y Seppälä consideran como posible objeción que la expresión “en un sentido” induciría el asentimiento a juicios aparentemente contradictorios porque sugeriría llevar a cabo una evaluación pragmática en lugar de evaluar las afirmaciones en función de sus valores de verdad. Y responden a esta objeción señalando que al evaluar los ejemplos de control los individuos no mostraron asentimiento a juicios aparentemente contradictorios como “En un sentido los leones son animales” y “En un sentido los leones no son animales”. Ahora bien, estas respuestas no permiten inferir si los individuos estaban efectuando una evaluación pragmática o de valores de verdad. Por más que la expresión “en un sentido” sugiera una evaluación pragmática de los juicios, esto no implica además que esta evaluación pragmática inducirá el asentimiento a cualquier par de juicios. Esto es así porque no todas las afirmaciones son aceptables, ni siquiera en un sentido pragmático. Simplemente, no todas las afirmaciones son pragmáticamente adecuadas. Esto parece evidente.

Machery y Seppälä apelan a las respuestas de los ejemplos de control para sostener que el uso de “en un sentido” no es lo que motiva el asentimiento a juicios aparentemente contradictorios. Por supuesto que no, pero esto no responde la objeción en cuestión. El problema es que las respuestas de los ejemplos de control no indican si esos juicios se están evaluando por sus condiciones de verdad o por su adecuación pragmática. Machery y Seppälä pretenden inferir, a partir de las respuestas de estos casos, que los individuos están efectivamente evaluando el valor de verdad de las afirmaciones. Pero, puesto que no hay un sentido pragmáticamente adecuado en que, por ejemplo, los leones no son animales, el hecho de que los individuos no acordaran con este juicio no sugiere que estuvieran evaluándolo por sus condiciones de verdad, puesto que cualquiera fuera el tipo de evaluación efectuada, la respuesta sería la misma. En este sentido, podría cuestionarse si las oraciones de control propuestas en el estudio son adecuadas. Los pares de control están confeccionados de tal modo que den lugar al mismo juicio sin importar si se asocia un prototipo o una teoría a los predicados en cuestión. Así, tanto si se tienen en cuenta las características típicas de los leones y de los animales, como si se tienen en cuenta las estructuras causales de rasgos o las

esencias de los leones y de los animales, se debería acordar con “En un sentido los leones son animales” y no con “En un sentido los leones no son animales”. Para este fin, las oraciones de control parecen funcionar. El problema es que este tipo de oración de control no permite extraer las conclusiones que luego pretenden extraer Machery y Seppälä, dado que no permite discernir si los individuos están evaluando la adecuación pragmática o las condiciones de verdad.

Por otra parte, Machery y Seppälä no tienen en cuenta otra objeción posible relacionada con el uso de “en un sentido”. A mi entender, más allá de si “en un sentido” induce una evaluación pragmática de las afirmaciones o no, el problema es que esta expresión hace que las afirmaciones no resulten contradictorias. Machery y Seppälä parecen tener en mente una inconsistencia lógica entre afirmaciones como las de los ejemplos y reconocen que la inconsistencia entre los juicios es sólo aparente. Sin embargo, se refieren de todos modos a ellos como juicios conflictivos, y esto no parece adecuado. Si se tomaran como verdaderas a la vez “Los tomates son verduras” y “Los tomates no son verduras” se estaría en presencia de una inconsistencia, por lo menos aparente. Pero considero que en el caso de “En un sentido, los tomates son verduras” y “En un sentido, los tomates no son verduras” la inconsistencia no es ni siquiera aparente. Estos juicios resultarían conflictivos sólo si se agregara que el sentido en ambas oraciones es el mismo, pero no hay nada en el estudio que sugiera esto.

Más aun, las aclaraciones que dejaron los participantes del estudio se pueden interpretar como intentos por explicitar en qué sentido estaban entendiendo los términos “tomate” o “verdura” en uno y otro caso, estableciendo diferencias claras entre ellos. Estos comentarios reflejan que los participantes estaban usando distintos tipos de información acerca de los tomates para evaluar los juicios, y, en este sentido, pueden servir como apoyo a cualquier propuesta que adopte una multiplicidad de estructuras conceptuales. Pero esto no implica necesariamente que cada estructura que representa tomates constituya un concepto independiente, sino que es compatible con que haya un único concepto TOMATE con distintas partes que contienen distintos tipos de información acerca de los tomates.

De modo que aun en el caso en que las teorías híbridas adoptaran la condición de coordinación, el estudio citado no constituiría evidencia en su contra, puesto que la contradicción entre los juicios es, según Machery y Seppälä sólo aparente o, como intenté defender, ni siquiera aparente. En síntesis, por un lado, las teorías híbridas no parecen aceptar la condición de coordinación. Más aún, adoptarla como criterio para determinar la independencia de conceptos impediría tratar a los conjuntos de ejemplares como conceptos, lo cual podría incluso resultar problemático para la propuesta pluralista del propio Machery,

que sostiene que cada conjunto de ejemplares de una categoría constituye un concepto. Y, por otro lado, aunque las teorías híbridas aceptaran la condición de coordinación, la evidencia de Machery y Seppälä contra esta condición es cuestionable. De manera que esta condición no permite establecer una buena distinción entre las teorías híbridas y pluralistas. En la próxima sección intentaré mostrar que la condición de asociación tampoco resulta adecuada.

2.2 La condición de asociación

Según la condición de asociación, las distintas partes de los conceptos están necesariamente conectadas o asociadas, de modo que “cuando una de las partes es usada, digamos, para categorizar, podemos *ipso facto* usar las otras partes del concepto para otros propósitos; por ejemplo, podemos usarlas para razonar deductiva o inductivamente” (Machery 2009, p. 64). Esta condición ha generado menos resistencia que la de coordinación, probablemente porque no parece ser la que presenta dificultades para las teorías híbridas. Si bien en principio permite distinguir entre enfoques híbridos y pluralistas, la adopción o el rechazo de esta condición no parece dar lugar a predicciones distintas que permitirían obtener evidencia empírica a favor o en contra de alguno de los enfoques. El hecho de que las distintas partes de un concepto tengan que ser accesibles o estar disponibles para el uso no implica que sean en efecto usadas siempre que una parte de un concepto se active, ni que deban activarse todas a la vez. A lo sumo, tal como interpreto esta condición, podría implicar la facilitación de la activación de otras partes del concepto al activar una. Si se entiende esta condición de este modo, no permite establecer distintas predicciones para las teorías híbridas y pluralistas puesto que, aun si se elaborara un estudio que mostrara, por ejemplo, la facilitación de representaciones teóricas de PERRO al activar ejemplares particulares de PERRO, esto podría explicarse tanto por una asociación entre partes del concepto PERRO como por asociaciones o relaciones entre conceptos independientes. Puede ser por esta razón que Machery concentra su discusión de los enfoques híbridos en las consecuencias empíricas que deberían seguirse de la condición de coordinación, que hemos visto en §2.1.

Por otra parte, a diferencia de la condición de coordinación, la condición de asociación es un supuesto que las teorías híbridas están dispuestas a aceptar, a pesar de que no todas lo formulen explícitamente. Es más, ésta resulta ser una condición que las teorías híbridas están dispuestas a defender, señalando las consecuencias que tendría su abandono. Sin embargo, la defensa de esta condición que llevan a cabo, por ejemplo, Margolis y Laurence (2010), sugiere que no la interpretan exactamente del mismo modo que Machery. Machery (2009) afirma que hay múltiples asociaciones entre conceptos y que los conceptos

correferenciales están usualmente asociados pero que, mientras que las asociaciones entre conceptos independientes son contingentes, las asociaciones entre las distintas partes de un concepto son necesarias. Sin embargo, como mencioné, no queda claro qué diferencias habría de hecho en la activación de los conceptos en uno u otro caso.

Machery parece atribuir a las teorías híbridas el supuesto de que la activación de una parte de un concepto necesariamente ocasiona la activación del resto. Por ejemplo, una teoría híbrida que postule que los conceptos tienen una parte prototípica y otra definicional sostendría que “necesariamente, el prototipo de x se puede usar en nuestros procesos cognitivos cuando ha sido usada la definición de x , y viceversa” (Machery 2009, p. 68). Denominaré a esta interpretación una versión *arquitectural* de la condición de asociación, según la cual, las partes de un concepto deben necesariamente activarse todas a la vez, o “estar listas para el uso” todas a la vez. Sin embargo, Machery no parece referirse simplemente a que deben “estar listas para el uso” a la vez, dado que distintos conceptos independientes también pueden estar listos a la vez para el acceso. Machery parece asumir que, al activar una parte de un concepto, el resto de las partes asociadas se activa también. Por esta razón denomino *arquitectural* a esta interpretación de la condición de asociación, porque parece describir a las partes de un concepto como partes de un bloque, como si estuvieran unidas estructuralmente y no pudieran funcionar independientemente unas de otras.

Ahora bien, las teorías híbridas no postulan la activación o uso simultáneo de todas las partes de un concepto. Por ejemplo, Osherson y Smith (1981), que proponen que los conceptos tienen una parte definicional y otra prototípica, sostienen que para algunos procesos, como la combinación de conceptos, utilizamos sólo la parte definicional. Margolis y Laurence (2007, 2010), que sostienen que los símbolos conceptuales están asociados a una estructura de procesamiento (i.e. aquella sobre la que operan los procesos, por ejemplo, de categorización) que incluye prototipos, teorías y ejemplares, sostienen que estas distintas partes de la estructura pueden usarse independientemente en los procesos de categorización. El modelo ACT-R de categorización, elaborado por Anderson y Betz (2001) propone estrategias de categorización que recurren o bien a una definición (una regla) o bien a una comparación con un conjunto de ejemplares. En este caso, el modelo “elige” mediante una función que “negocia” entre la confianza en el resultado y el tiempo que le llevará arribar a él, si optará por la vía definicional o la comparación con los ejemplares. En este modelo, las dos vías son independientes. Así, ninguna de estas teorías sostiene que necesariamente todas las partes de un concepto deban activarse simultáneamente. En este sentido las teorías

híbridas generalmente rechazan la condición de asociación entendida en sentido arquitectural.

Sin embargo, las teorías híbridas adhieren a la condición de asociación, entendida en un sentido *semántico*. Dos buenos ejemplos de este modo de interpretar la condición son las propuestas de Prinz (2002), que vimos en el capítulo 5, y Margolis y Laurence (2007, 2010). La propuesta de Prinz puede leerse como una propuesta híbrida, puesto que considera que los conceptos tienen partes heterogéneas: incluyen prototipos, teorías y conjuntos de ejemplares. Prinz (2002, cap. 6) señala que, por cuestiones arquitecturales relacionadas con la capacidad de la memoria de trabajo, sería imposible activar simultáneamente todas las partes de un concepto. Por eso, como detallé en el capítulo 5, caracteriza los conceptos como los conjuntos de representaciones que pueden activarse en la memoria de trabajo para representar una categoría. Distintos conjuntos de representaciones pueden activarse para representar la misma categoría en distintos contextos. Prinz en muchas oportunidades se refiere a estos distintos conjuntos de representaciones como partes del mismo concepto. Ahora bien, el sentido en el que los considera partes del mismo concepto puede entenderse como un sentido semántico, como detallaré a continuación.

Como se vio en el capítulo 5, Prinz propone que los conceptos tienen dos tipos de contenido: un contenido intencional o real y un contenido cognitivo o nominal. El contenido real refiere a esencias reales de las cosas y el contenido nominal, en cambio, refiere a las propiedades a partir de las cuales identificamos las instancias en el mundo que caen bajo un concepto. Los distintos conjuntos de representaciones que se activan en cada ocasión para representar una determinada categoría son correferenciales, puesto que comparten un contenido real. Pero, a la vez, tienen distintos contenidos cognitivos o nominales, dado que sirven para detectar y representar distintas propiedades de apariencia. De modo que permiten explicar los juicios aparentemente contradictorios tratados en la sección anterior, puesto que en un sentido comparten el contenido y en otro sentido no. Los conceptos correferenciales pueden activarse independientemente uno del otro, pero están asociados en redes de memoria a largo plazo como representaciones de una misma categoría. En este último sentido puede decirse que son partes de un mismo concepto. Tomando el ejemplo de tres conjuntos de representaciones perceptivas que sirven para detectar y representar *whisky*, Prinz afirma:

Cada uno de estos tres proxitipos tiene un contenido nominal distinto. De todos modos se puede pensar en todos ellos como partes de un mismo continuo, basado en experiencias con la misma sustancia (...). Este continuo consta de proxitipos, pero esos proxitipos están conectados por su

colaboración en detectar whisky (...). Whisky es el contenido real de tal continuo, y de los proxitipos que lo componen. (2002, p. 280)

En la propuesta de Margolis y Laurence (2007, 2010) los conceptos son símbolos semejantes a palabras del lenguaje natural, como los símbolos de un LDP, que tienen una estructura semántica y una estructura para el procesamiento. La estructura para el procesamiento puede contener prototipos, ejemplares y teorías, que están asociados a los símbolos. Adoptando una propuesta de tipo fodoriano, sostienen que los símbolos se individualizan por su contenido más ciertas propiedades del vehículo que permiten distinguir entre conceptos correferenciales que difieren en contenido cognitivo (Margolis & Laurence 2007). En esta propuesta, un prototipo y una teoría se pueden considerar correferenciales si están asociados al mismo símbolo. En este sentido adhieren a la condición de asociación. Sin embargo, explícitamente rechazan que la activación de una parte de la estructura para el procesamiento deba necesariamente activar otras partes.

Margolis y Laurence (2010) señalan que, si se rechaza la condición de asociación, el pluralismo de conceptos tiene dificultades para justificar la correferencialidad de los distintos conceptos. En una propuesta pluralista se asume que, por lo menos para algunas categorías, disponemos de distintas representaciones correferenciales e independientes. Pero, ¿cómo se justifica la correferencialidad si el conjunto de elementos que caen bajo el prototipo de, por ejemplo, PEZ, puede incluir ballenas, mientras que éstas no caen bajo la teoría de PEZ? Según Margolis y Laurence, la asociación de las distintas estructuras con una única representación semántica que fija la referencia del concepto sería lo que garantiza la correferencialidad. En este sentido, las teorías híbridas presentarían una ventaja sobre las pluralistas.

Sin embargo, considero que conceder esta conclusión a Margolis y Laurence sería demasiado apresurado. Antes de conceder esa conclusión, se debería intentar explicitar qué compromisos asumen las teorías pluralistas respecto de las asociaciones entre los conceptos. En particular, es necesario aclarar qué tesis presuponen afirmaciones como que distintas estructuras conceptuales pueden ser correferenciales y si esta correferencia implica algún tipo de relación especial entre tales estructuras. Con respecto a la correferencia, Machery señala que ésta podría estar garantizada por una semántica informacional y recurre, precisamente, a la propuesta de Prinz para ilustrar cómo podría funcionar. Sugerir que una propuesta como la de Prinz le podría servir al pluralismo apoya mi intuición de que las teorías pluralistas también adhieren a la condición de asociación cuando ésta se interpreta en sentido semántico.

Esto es así porque, como señalé más arriba, la propuesta de Prinz explícitamente señala en qué sentido los prototipos, ejemplares y teorías correferenciales forman parte de un mismo concepto y este, como vimos, es un sentido semántico.

De hecho, hay propuestas pluralistas que, si bien sostienen que cada categoría está representada por diversos conceptos independientes, incorporan asociaciones específicas entre conceptos correferenciales, agrupando de un modo especial a todos los conceptos que representan una misma categoría. Tal es el caso de la propuesta de Weiskopf (2009a, 2009b) que se presenta como pluralista, y trata a los prototipos, ejemplares y teorías como conceptos independientes, pero sin embargo incorpora archivos mentales para agrupar los conceptos correferenciales y así explicar los distintos fenómenos relacionados con el contenido. Si bien puede haber múltiples asociaciones entre conceptos, hay un tipo de asociación específica para los conceptos correferenciales, a la que Weiskopf denomina archivos mentales. La necesidad de establecer una asociación específica entre conceptos correferenciales tiene que ver con cuestiones semánticas a las que las teorías de conceptos deben responder. La mención de archivos mentales (Prinz 2002, Weiskopf 2009b) o símbolos (Margolis y Laurence 2007) son distintos modos de asociar estructuras diversas en función de la individuación semántica de los conceptos.

Otras consideraciones también sugieren que el pluralismo adhiere a la condición de asociación semántica. En particular, si se sostiene que los conceptos correferenciales son completamente independientes entre sí, ¿esto implica que la relación entre $TOMATE_P$ y $TOMATE_T$ es del mismo tipo que la que hay entre $TOMATE_P$ y $REPOLLO_P$? Esto implicaría que usamos los distintos tipos de conceptos que refieren a tomates como si no fueran correferenciales. De modo que parece ser que el pluralismo trata todos los casos de conceptos correferenciales e independientes como casos Frege, i.e. como casos de conceptos coextensivos que tienen poderes causales distintos y en los que el individuo no sabe que ambos refieren a la misma propiedad en el mundo. Tratar de este modo al prototipo y al conjunto de ejemplares que representan la clase de los tomates resultaría sumamente antiintuitivo. La relación entre $TOMATE_P$ y $TOMATE_T$ no parece ser del mismo tipo que la que hay entre, e.g. $SUPERMAN$ y $CLARK KENT$.⁴⁹ Pero el pluralismo no dispone de recursos para negar esa simetría, a no ser que adopte la condición de asociación.

La objeción de Margolis y Laurence en relación a la correferencia presupone que una teoría adecuada de conceptos debe dar cuenta, entre otros requisitos, del contenido

⁴⁹ No quisiera comprometerme aquí con algún tipo de estructura semántica interna en particular para los conceptos de individuos, por eso no incluí especificaciones al respecto. Pueden leerse como etiquetas atómicas.

conceptual, debe explicar en qué consiste que un concepto sea acerca de algo. Como señalé en el capítulo 1, generalmente se engloban en el requisito de contenido tres aspectos: qué tipo de información portan los conceptos (por ejemplo, si es conocimiento proposicional, estadístico, o de otro tipo y si la información está estructurada o es atómica), qué tipo de relación tienen con el mundo (por ejemplo, si se trata de una relación causal o de similitud), y qué relación tienen con los otros conceptos. Sin embargo, no todos los autores consideran que el contenido sea un requisito fundamental. Como se vio en el capítulo 1, Machery (2009) distingue entre las teorías psicológicas y filosóficas de conceptos precisamente en función de los requisitos explicativos que éstas adoptan y, al desarrollar las características de los conceptos, tiene en cuenta sólo los requisitos explicativos típicamente asociados a las teorías psicológicas. Sin embargo, como pone en evidencia la crítica de Margolis y Laurence, a pesar de no ocuparse de los *desiderata* que considera filosóficos, Machery presupone soluciones a los mismos, especialmente al requisito de contenido. La mención de conceptos correferenciales pero independientes es un ejemplo de esto. De modo que la objeción de Margolis y Laurence, lejos de exigir respuestas a problemas que no estaban dentro de los objetivos explicativos del pluralismo de conceptos, tales como el contenido, permite explorar qué compromisos presupone una propuesta pluralista en relación a este requisito. Y el análisis que presenté sugiere que, al igual que el enfoque híbrido, el enfoque pluralista también presupone una condición de asociación en sentido semántico.

En esta sección sostuve que las teorías generalmente consideradas híbridas no se comprometen con la condición de coordinación y si bien aceptan la condición de asociación, no lo hacen en el sentido que Machery cuestiona. Más aún, el tratamiento (o la presunción) de la correferencialidad en las propuestas pluralistas y la pretensión de que la relación entre TOMATE_P y TOMATE_T no sea del mismo tipo que la relación entre, por ejemplo, SUPERMAN y CLARK KENT, sugiere que una condición similar a la de asociación en sentido semántico puede estar implícita también en las teorías pluralistas. En la próxima sección discutiré un dilema planteado por Weiskopf (2009a) contra el enfoque híbrido. Sostendré que al igual que los argumentos de Machery, se basa en una lectura arquitectural de la noción de “parte de un concepto” que no es la que las teorías híbridas parecen proponer. Asimismo, señalaré que, si se lee “parte de un concepto” de un modo semántico y no arquitectural, las objeciones propuestas por Weiskopf tienen el mismo peso contra las teorías híbridas que contra las pluralistas.

3. El dilema de Weiskopf

Weiskopf (2009a) descarta las teorías híbridas a partir de un dilema. Sostiene que, o bien incorporan algunos tipos de estructuras conceptuales y entonces presentan ciertas ventajas sobre las teorías que adoptan un único tipo de estructura, pero no consiguen dar cuenta de todos los fenómenos, o bien incorporan todos los tipos de estructuras necesarios para dar cuenta de los efectos empíricos observados y entonces se tornan inmanejables. No tengo objeciones al primer cuerno del dilema. Sin embargo, se puede señalar que el blanco de este cuerno son sólo las primeras propuestas híbridas, como la de Osherson y Smith (1981) que incorpora sólo definiciones y prototipos, o el modelo RULEX (por *rule-plus-exception*, reglas más excepciones) de Nosofsky y colegas (1994) que incorpora sólo definiciones y ejemplares. Las teorías contemporáneas, que acuerdan en incorporar por lo menos prototipos, ejemplares y teorías, y consideran incluso la posibilidad de que la evidencia empírica lleve a la incorporación de otras estructuras, son blanco del segundo cuerno.

Ahora bien, ¿en qué sentido serían inmanejables los conceptos propuestos por estas teorías? En este punto parece que Weiskopf, al igual que Machery, atribuye a las teorías híbridas una condición de asociación en sentido arquitectural. Si se pretende que todas las estructuras que son parte de un concepto deban ser activadas y procesadas en simultáneo, entonces los conceptos híbridos así entendidos pueden resultar inmanejables. Esto se debe a las limitaciones de la memoria de trabajo, mencionadas más arriba. Sin embargo, como vimos, las teorías híbridas no adhieren a la condición de asociación en este sentido. Y sin suponer esta condición, la objeción pierde sentido, puesto que no hay otro modo de interpretar la crítica de “no ser manejables” que no sea a la vez una objeción al propio pluralismo. Las teorías híbridas y las pluralistas coinciden en postular una multiplicidad de representaciones correferenciales que pueden usarse de modo independiente en los procesos cognitivos, de manera que ambos enfoques deben incorporar mecanismos que determinen qué representación se activa en un contexto dado o cómo intervienen distintos tipos de representaciones para llevar a cabo un determinado proceso. Si el pluralismo puede postular tales mecanismos y “manejar” la multiplicidad de representaciones en este sentido, no hay razones para suponer que los enfoques híbridos no puedan hacer lo mismo. La articulación y organización de las diversas representaciones correferenciales es algo que deben resolver tanto una propuesta de estructuras independientes como una que proponga estructuras que son partes (en sentido semántico) de un mismo concepto.

Weiskopf señala que muchos modelos híbridos proponen que no deben activarse todas las partes de un concepto simultáneamente y que cada parte puede responder selectivamente a

distintos estímulos o distintas tareas. Este sería el caso del modelo ACT-R, de Anderson y Betz (2001) por ejemplo. Este tipo de propuestas son, según Weiskopf, indistinguibles de las propuestas pluralistas. De modo que, evidentemente, Weiskopf está trazando la distinción entre los enfoques híbridos y pluralistas en función de la activación independiente o no de las distintas estructuras. Si las teorías híbridas no defienden una asociación arquitectural, entonces no son blanco de esta objeción.

¿Esto convierte a la distinción entre teorías híbridas y pluralistas en una cuestión meramente terminológica? En la sección siguiente sostendré que sí y sugeriré algunos motivos por los cuales igualmente siguen utilizándose estos términos distintos.

4. Algunas estrategias para rechazar el eliminativismo

Recordemos el argumento general de Machery (2009, 2010, Machery & Seppälä 2011) contra el enfoque híbrido: si las teorías híbridas rechazan las condiciones de coordinación y asociación (y no proponen otras que permitan distinguir entre partes de conceptos y conceptos independientes), entonces nada las diferencia de las pluralistas. Y, alternativamente, si aceptan esas condiciones, entonces hay evidencia suficiente en su contra como para abandonarlas. En §2.1 discutí la segunda alternativa y sostuve que la evidencia contra la coordinación es muy débil. Ahora bien, a lo largo del capítulo defendí que las teorías híbridas no adhieren a las dos condiciones que propone Machery. Y si bien adhieren en un sentido a la condición de asociación, lo hacen de un modo en que parecen coincidir con el enfoque pluralista. Esto sugiere que las propuestas híbridas y pluralistas contemporáneas comparten sus tesis centrales. Éstas coinciden en postular una multiplicidad de representaciones correferenciales heterogéneas que se pueden activar independientemente unas de otras en los procesos cognitivos. Asimismo, postulan un nivel semántico de asociación entre las representaciones correferenciales.

Una cuestión que parece estar en juego al presentar los enfoques híbridos y pluralistas como opuestos es en qué nivel deben individuarse los conceptos. A simple vista, parece ser que los enfoques híbridos identifican los conceptos con la unión de todas las representaciones de una categoría, mientras que los pluralistas identifican los conceptos con cada una de las representaciones de una categoría que pueden activarse independientemente. Así, parecería que los enfoques híbridos adhieren al principio de la singularidad, i.e. la tesis de que cada categoría está representada por un concepto (Weiskopf 2009a), mientras que los enfoques híbridos la abandonan y postulan en cambio una multiplicidad de conceptos para cada categoría. Si bien en la caracterización de qué entienden por concepto uno y otro enfoque,

éste parece ser el caso, en las secciones previas intenté mostrar que cuando se analizan en detalle los compromisos arquitecturales y semánticos de ambos enfoques, la diferencia con respecto a la tesis de la singularidad es meramente terminológica, como sintetizaré a continuación.

Como vimos, las teorías híbridas no tratan las representaciones asociadas como un bloque que se activa siempre en conjunto, sino que, al igual que las teorías pluralistas, afirman que distintas representaciones pueden funcionar de modo independiente. En este sentido, admiten que hay distintos tipos de representaciones que pueden ser utilizadas de modo independiente para pensar acerca de una categoría y coinciden, en consecuencia, en asociar una multiplicidad de estructuras independientes con cada categoría. Así, en un sentido arquitectural no adhieren al principio de la singularidad. Por otro lado, las teorías pluralistas también incorporan asociaciones entre las representaciones de una misma categoría, al igual que las teorías híbridas. De este modo, ambos enfoques apelan a cierta singularidad semántica, en tanto que para dar cuenta de ciertos fenómenos semánticos postulan un nivel de individuación de conceptos por categorías, y no por estructura interna. De modo que esta aparente individuación de conceptos en distintos niveles resulta ser más una diferencia de énfasis que una diferencia en tesis centrales y ambos enfoques tienen en cuenta tanto el nivel de activación en la memoria de trabajo como el nivel de asociación semántica a largo plazo.⁵⁰

¿Por qué hay tanta resistencia, en especial por parte de los defensores de las llamadas teorías híbridas, a aceptar que ambos enfoques están de acuerdo en sus tesis principales? Considero que parte de esta resistencia puede deberse a la asociación que propone Machery entre pluralismo y eliminativismo. Machery defiende la que él denomina hipótesis de la heterogeneidad, según la cual cada categoría está representada por varios conceptos de distinto tipo. En este sentido, la hipótesis de la heterogeneidad forma parte de las tesis pluralistas. Sin embargo, Machery rechaza el pluralismo y defiende un eliminativismo de la noción de “concepto”. Según Machery (2005, 2009) las estructuras conceptuales son tan heterogéneas que no comparten suficientes propiedades relevantes como para constituir una clase natural y, por lo tanto, el término “concepto” no juega ningún papel en el avance científico o, peor aún, resulta perjudicial, puesto que limita la investigación de cada una de las estructuras que suelen agruparse en la clase de los conceptos y, en consecuencia, debe ser

⁵⁰ Como sostuve en relación al neo-empirismo de conceptos, la apelación a dos niveles distintos para la individuación de conceptos, i.e. por un lado el nivel semántico que identifica distintos tipos de estructuras como representaciones de una misma categoría y, por otro lado, el nivel de las distintas estructuras que pueden activarse independientemente en la memoria de trabajo, acarrea ciertas dificultades relativas al contenido, que cualquier propuesta pluralista deberá enfrentar. En el próximo capítulo ofreceré una respuesta posible a estas dificultades.

eliminado del vocabulario científico y reemplazado por “prototipos”, “ejemplares” y “teorías”.

La estrategia eliminativista de Machery puede sintetizarse como una estrategia de tipo “divide y reinarás”. En primer lugar, como vimos en el capítulo 1, distingue entre las teorías psicológicas y filosóficas de conceptos en función de los requisitos de los que cada una se ocupa, y descarta las teorías filosóficas, por considerar que se ocupan de fenómenos relacionados con la metafísica de los pensamientos, que no tienen relación con aquellos que ocupan a los psicólogos que llevan a cabo investigaciones empíricas de la estructura conceptual. Así, señala que su interés está en la noción de “concepto” como se usa en las teorías psicológicas. De este modo, sienta las bases para descartar características que los distintos tipos de estructuras conceptuales podrían tener en común, en particular, las relacionadas con el contenido, y con la combinación de conceptos, que presumiblemente deba operar sobre los distintos tipos de representaciones conceptuales disponibles. Como argumenté en el capítulo 1, esta división no recoge los objetivos explicativos de muchos de los investigadores que trabajan con conceptos, parece ignorar los aspectos interdisciplinarios de la investigación en ciencias cognitivas y, por último, no tiene en cuenta que por más que algunos de los requisitos no sean explícitamente abordados por ciertos enfoques, algunas de las tesis que sostienen presuponen determinadas respuestas a tales requisitos. Por estos motivos, la estrategia que propongo para rechazar la tesis eliminativista empieza por cuestionar el primer paso de la argumentación de Machery, i.e. empieza por rechazar la distinción entre teorías psicológicas y filosóficas de conceptos.

Ahora bien, retomando la estrategia argumentativa de Machery, una vez establecida la distinción entre teorías psicológicas y filosóficas, y establecida la prioridad de los requisitos típicamente psicológicos, Machery distingue entre teorías híbridas y pluralistas, con el objetivo de establecer la preferencia de las teorías pluralistas por sobre las híbridas. Así, el dilema propuesto a las teorías híbridas señala que, o bien se diferencian de las pluralistas adoptando las condiciones de asociación y coordinación, y entonces deben ser abandonadas por la evidencia en su contra; o bien rechazan las condiciones y entonces no se diferencian de las pluralistas, quedando sujetas a la misma suerte que estas últimas. Por último, a partir del pluralismo de estructuras conceptuales y su heterogeneidad, Machery defiende el eliminativismo de la noción de “concepto”.

Pues bien, esta estrategia no pasa desapercibida a los defensores de teorías que incorporan varios tipos de estructuras. Considero que parte de la defensa de las teorías híbridas está orientada a rechazar la conclusión eliminativista, apelando a ciertas propiedades

en común de todos los conceptos.⁵¹ Algunos de los argumentos de los defensores de las teorías híbridas señalan que la estructura híbrida común a la mayoría de conceptos permite estudiarlos como una clase de interés científico. Este es un modo de bloquear la conclusión eliminativista. Sin embargo, hay otras estrategias más fructíferas que la defensa de las tesis híbridas para rechazar el eliminativismo, que desarrollaré a continuación.

Hasta ahora presenté argumentos que objetan los primeros pasos de la estrategia eliminativista de Machery. Ofrecí razones para rechazar la distinción entre teorías psicológicas y filosóficas de conceptos. Estas razones acompañan la intención de tener en cuenta requisitos explicativos tradicionalmente tratados por los filósofos que, al incorporarse a las teorías de conceptos, revelan tesis compartidas por los enfoques híbridos y pluralistas, en particular ciertos fenómenos asociados al contenido conceptual. Sostuve que la distinción entre teorías híbridas y pluralistas por medio de las condiciones de coordinación y asociación no resulta adecuada y que las razones que ofrece Machery para preferir las teorías pluralistas por sobre las híbridas no son suficientes, puesto que se basan en una distinción que no es adecuada y en evidencia sumamente débil. Ahora bien, se requieren argumentos que permitan discutir el último paso de la estrategia eliminativista que pasa de la postulación de un conjunto de entidades heterogéneas, a la conclusión que sugiere eliminar el término “concepto”. Si bien al tener en cuenta los requisitos de contenido y composicionalidad se revelan aspectos comunes a todos los conceptos, es innegable que los distintos tipos de estructuras conforman una clase heterogénea, y el argumento eliminativista descansa fuertemente en este hecho.

Machery señala que dada la heterogeneidad de las estructuras que se agrupan bajo el término “concepto”, el término no recoge una clase natural. “Si ‘concepto’ no recoge una clase natural” sostiene Machery, “entonces es poco probable que resulte una noción útil en psicología. Incluso es probable que se interponga en el camino del progreso en psicología impidiendo el desarrollo de un esquema clasificatorio que identificaría las clases naturales relevantes” (2009, p. 230). Machery sostiene que reemplazar el término “concepto” por, “prototipo”, “ejemplar” y “teoría”, permitiría a los científicos investigar adecuadamente cada tipo de entidad, en lugar de buscar generalizaciones que se apliquen a todas ellas. Este argumento supone que las únicas clases sobre las que pueden realizarse generalizaciones de interés científico son las clases naturales. Presentaré una serie de objeciones a este supuesto. Primero, señalaré cómo se delimita usualmente la clase de los conceptos, segundo,

⁵¹ Véase, por ejemplo, Gonnerman & Weinberg (2010), Keil (2010), Margolis & Laurence (2010), Scarantino (2010).

reconstruiré la noción de “clase natural” que adopta Machery y, tercero, señalaré por qué si bien los conceptos no se adecúan a tal noción, pueden constituir una clase de interés científico.

Generalmente, el criterio para determinar qué tipos de entidades son los conceptos es un criterio que apela al rol funcional. Se establecen una serie de roles que los conceptos deben satisfacer en nuestra vida mental y se postula luego cuál es la entidad que mejor satisface todos esos roles. Las diferencias entre las entidades que postula cada teoría tienen que ver en algunos casos con los roles explicativos seleccionados. Así, por ejemplo Fodor (1999) establece que los conceptos son las clases de cosas que funcionan como causas y efectos de los procesos mentales, permiten explicar gracias a su composicionalidad la productividad y sistematicidad del pensamiento, son utilizados como categorías y son públicos. En este sentido, establece una serie de funciones que los conceptos deben cumplir. Los argumentos que conectan luego estos roles con el LDP pretenden mostrar que lo único que puede satisfacer todas esas funciones de manera conjunta son las representaciones mentales simbólicas estructuradas a la manera de los lenguajes naturales. De modo que arriba así a una caracterización ontológica de los conceptos: estos son símbolos de un LDP que son particulares mentales físicamente instanciados.

Prinz también adopta un criterio funcional, pero focaliza el rol de los conceptos en la categorización. A pesar de que establece una extensa lista de *desiderata* para una teoría de conceptos, el criterio que predomina para diferenciar conceptos de otras entidades es el de ser utilizados en la detección de categorías, junto con la condición de estar bajo control endógeno. Este también es un criterio de tipo funcional, puesto que establece los roles que los conceptos deben cumplir: deben funcionar como mecanismos detectores de categorías y deben estar bajo control del organismo. Ahora bien, para satisfacer estos roles Prinz también postula particulares mentales, pero de un tipo diferente al que propone Fodor: los proxitipos que, como vimos en el capítulo 5, consisten en copias de representaciones perceptivas. El pluralismo de Weiskopf (2009a), a su vez, incluye como roles explicativos de los conceptos el hecho de que son utilizados para categorizar y que se pueden combinar de manera productiva y sistemática en estructuras más extensas y complejas. El eliminativismo de Machery (2009), por otra parte, sostiene que los conceptos son los que se usan por *default* en procesos cognitivos superiores y se incluyen entre estos procesos la categorización, la deducción y la inducción.

De este modo los distintos autores fijan un criterio para delimitar la clase de los conceptos, i.e. para determinar qué entidades serán conceptos. Una cuestión diferente, pero

relacionada con esta, es si los conceptos constituyen una clase natural. Una vez establecidos los roles funcionales que los conceptos deben satisfacer y una vez halladas las entidades que los satisfacen, ¿hemos recortado así una clase natural o hemos delimitado de esta manera un conjunto de entidades mentales análogo a ‘cosas que hay en la cocina’, el cual no constituye una clase natural? Es fundamental observar que la cuestión del criterio para seleccionar conceptos y la pregunta por si los conceptos constituyen una clase natural son distintas. Todos los autores comienzan por establecer una serie de roles funcionales que los conceptos deben satisfacer y algunos, como Fodor y Prinz, defienden luego que las entidades que ocupan o satisfacen ese rol constituyen una clase natural.

Pues bien, en relación a la pregunta por si los conceptos constituyen una clase natural, en este caso la cuestión importante es si las entidades agrupadas comparten un conjunto de propiedades relevantes en común, de modo que se puedan establecer generalizaciones fructíferas para la ciencia sobre todo ese conjunto (Boyd 1991, Machery 2009). En relación a esto, los autores que adhieren a un eliminativismo sostienen que las entidades que satisfacen los roles funcionales de los conceptos no comparten propiedades relevantes, por lo que no constituyen una clase natural y, por lo tanto, el término “concepto” debe ser eliminado del vocabulario científico. En cambio, autores, como Fodor o Prinz, que sostienen que los conceptos son una clase natural postulan generalmente un único tipo de entidad con un conjunto de propiedades metafísicas en común y argumentan que por lo tanto están describiendo una clase natural. Entre estas dos posturas puede encontrarse, sin embargo, una tercera, que admite que las entidades que satisfacen esos roles funcionales no constituyen una clase natural, i.e. que no se pretende cortar así el mundo “por sus articulaciones”. No obstante, se pueden establecer generalizaciones fructíferas sobre esas entidades, de modo que conservar la noción de concepto resulta útil para la ciencia. Esta postura sostiene así que los conceptos constituyen una clase funcional y considero que es una respuesta adecuada al eliminativismo.

Así, puede concederse que los conceptos no constituyen una clase natural, sino una clase funcional, e igualmente sostener que resulta fructífero para la ciencia conservar el término, puesto que de este modo se pueden realizar generalizaciones sobre los distintos elementos de la clase concepto que resultan relevantes para explicar el funcionamiento de la mente (Lalumera 2010, Weiskopf 2009a, 2010). Weiskopf (2009a, 2010) sostiene que dados los roles que los conceptos deberían satisfacer, la única manera de dar cuenta de todos los fenómenos que los conceptos deberían explicar es apelando a un pluralismo de estructuras. Así, concuerda con Machery en que los conceptos no constituyen una clase natural, pero

sostiene, sin embargo, tal como se mencionó, que conservar el término “concepto” es útil para la ciencia, puesto que pueden establecerse generalizaciones fructíferas sobre esta clase de entidades.

Weiskopf sostiene que: (a) hay procesos mentales que dependen de la forma lógica y que operan sobre la clase de los conceptos en general, (b) la combinación conceptual opera sobre distintos tipos de estructuras, (c) los conceptos tienen modos comunes de adquisición y (d) los procesos que permiten el almacenamiento en la memoria a largo plazo y la posterior recuperación, manipulan distintos tipos de representaciones. Considero que estos son buenos ejemplos de generalizaciones fructíferas que alientan a rechazar un eliminativismo y a conservar el término “concepto” en la psicología. Sin embargo, señalaré una consecuencia de (a) que no es compatible con el tipo de pluralismo que defenderé y sugeriré una ligera modificación de (c). Me ocuparé de todas ellas a continuación, comenzando por la última.

Respecto de (d), dado que en principio el sistema de memoria a largo plazo en el que se almacenan los distintos tipos de representaciones es uno sólo, resulta necesario que los procesos que permiten almacenar y luego recuperar las representaciones sean capaces de manipular todos los tipos de estructuras conceptuales que somos capaces de procesar. Así, deberían poder formularse generalizaciones acerca del almacenamiento y la recuperación que abarquen a todos estos tipos de representaciones. En relación a (c), Weiskopf sostiene que es posible que dispongamos de mecanismos que nos permiten adquirir distintos tipos de conceptos. Como ejemplo, señala que es habitual explicar la adquisición de prototipos por abstracción a partir de representaciones de ejemplares, o la adquisición de teorías a partir de prototipos, utilizando distintos mecanismos mediados en general por el lenguaje natural. Sin embargo creo que, basándose en estos ejemplos, una forma más clara de expresar esta generalización sería decir que hay mecanismos de adquisición que manipulan distintos tipos de conceptos, lo cual, como sostendré en el capítulo 7, no significa que estos distintos conceptos se adquieran del mismo modo.

En relación a (b), Weiskopf ofrece una generalización interesante: es posible combinar la información de distintos conceptos para formar una representación compleja, de modo que el proceso de combinación conceptual debe poder recibir como *input* los distintos tipos de representaciones. De hecho, el propio Machery (2009, cap. 7) también sugiere cómo intervienen los distintos tipos de estructuras conceptuales en los procesos de combinación y apela, para esto, a un único proceso, que opera sobre distintos tipos de conceptos. Por último, respecto de (a), i.e. la forma lógica, Weiskopf señala que si los pensamientos tienen estructura lógica, los procesos que operan en virtud de ella deben funcionar independientemente del tipo

de representaciones específicas que los conforman. Por ejemplo, a partir del pensamiento de que el niño tiene sed y no hay naranjas, podemos inferir el pensamiento de que el niño tiene sed o el pensamiento de que no hay naranjas, independientemente de cómo representemos al niño sediento o a las naranjas. Sin embargo, como vimos en el capítulo 4, este tipo de operaciones, puramente dependientes de la forma sintáctica (y no de la estructura semántica) sólo pueden realizarse sobre representaciones simbólicas de tipo amodal. Así, un pluralismo amplio como el que desarrollaré en los próximos capítulos, que postula no sólo una multiplicidad de estructuras semánticas, sino también de vehículos, no admite esta última generalización, dado que ésta valdrá sólo para las representaciones amodales, pero no para las perceptivas.

Dado que se puede defender que se pueden formular generalizaciones de interés científico sobre la clase de los conceptos, la apelación a algún tipo de homogeneidad en las representaciones conceptuales (tal como sería afirmar que todas son representaciones híbridas que incluyen determinadas partes) no es la única opción para rechazar la conclusión eliminativista. Es posible defender la heterogeneidad de las representaciones conceptuales, sin por eso caer en la postulación de una clase que no resulta de utilidad científica. Por este motivo, considero que no es necesario buscar aspectos de uniformidad en las representaciones conceptuales. Como defendí en este capítulo, las propuestas usualmente llamadas híbridas y las usualmente llamadas pluralistas están de acuerdo en determinados puntos centrales. Ambos enfoques coinciden en postular una multiplicidad de representaciones conceptuales heterogéneas, respecto de las cuales postulan, por un lado, una independencia “arquitectural”, i.e. postulan que es posible activar estas estructuras por separado, de modo independiente y, por otro lado, una asociación de tipo semántico, i.e. postulan asociaciones en memoria a largo plazo para representaciones correferenciales que deben ser de un tipo distinto a las asociaciones con otras representaciones con las que no comparten la referencia. En este sentido, podría decirse que estas propuestas son pluralistas en sentido arquitectural e híbridas en sentido semántico.

Ahora bien, esto no quiere decir que las distintas propuestas acuerden en todas las tesis y que no haya diferencias relevantes entre ellas. Por supuesto que no es lo mismo incorporar, por ejemplo, sólo definiciones y prototipos como estructuras conceptuales que admitir también ejemplares y teorías. El debate fructífero entre las distintas teorías que incorporan una multiplicidad de estructuras conceptuales distintas, entonces, es el que se da en torno a modelos específicos, compromisos concretos con determinados tipos de estructuras y propuestas precisas sobre cómo están asociadas estas estructuras, así como de qué depende

que se active una u otra. Sin embargo, la mayor parte de la evidencia que se discute actualmente y que las distintas teorías de conceptos intentan acomodar, proviene de estudios que pretenden evaluar predicciones de determinadas propuestas de prototipos, ejemplares o teorías, que adhieren al principio de uniformidad. Por esta razón, para comparar y evaluar distintas teorías que incorporan una pluralidad de estructuras conceptuales, es necesario un mayor desarrollo de modelos que incorporen tal pluralidad de estructuras, cuyas predicciones específicas puedan ser puestas a prueba. Con este objetivo en mente, en los próximos capítulos ofreceré los lineamientos generales de una propuesta pluralista que incorpora no sólo distintos tipos de estructuras de información, sino también distintos tipos de vehículos conceptuales.

Capítulo 7

El pluralismo amplio

El análisis llevado a cabo en los capítulos 2 a 5 motiva la postulación de distintos tipos de estructuras conceptuales para dar cuenta de los requisitos fundamentales que debería satisfacer una teoría adecuada de conceptos: intencionalidad, alcance, composicionalidad, categorización adquisición y publicidad. Como sostuve, definiciones, prototipos, ejemplares, teorías y representaciones atómicas parecen adecuarse a distintos fenómenos y, en consecuencia, aquellas teorías que postulan estas diversas estructuras como parte de la clase de los conceptos parecen contar con una ventaja explicativa. Una apreciación similar puede sostenerse respecto de los vehículos conceptuales. Tanto los símbolos amodales como las representaciones perceptivas resultan adecuados para distintos fenómenos, de modo que proponer una articulación entre ambos supondría una ventaja. A partir de argumentos similares, algunas teorías de conceptos contemporáneas afirman que la clase de los conceptos es una clase heterogénea y que cada categoría puede estar representada por estructuras semánticas correferenciales de distinto tipo. Como argumenté en el capítulo 6, estas teorías parecen acordar en dos tesis centrales, que tomaré como punto de partida: (i) las distintas estructuras conceptuales correferenciales pueden ser activadas con independencia en la memoria de trabajo y (ii) estas estructuras mantienen en la memoria a largo plazo una asociación específica, distinta de la que pueden tener con otras representaciones.

En este capítulo ofreceré, a partir de estas dos tesis, los lineamientos principales de una propuesta pluralista: el pluralismo amplio. Señalaré las ventajas que ofrece esta propuesta en relación a los requisitos fundamentales y sugeriré cómo pueden enfrentarse los nuevos desafíos que el pluralismo acarrea, tales como explicar la relación entre los distintos tipos de estructuras de información y entre los distintos tipos de vehículos, qué tipos de mecanismos pueden operar sobre las distintas estructuras y cómo se determina la activación de una estructura u otra. En §1 reconstruiré el argumento a favor del pluralismo que atraviesa los capítulos 2 a 5. Dadas las limitaciones de cada tipo de estructura y de vehículo para dar cuenta de los fenómenos fundamentales que las teorías de conceptos se proponen explicar, y dadas sus ventajas para explicar algunos, sostendré que la mejor estrategia disponible consiste en postular una multiplicidad de estructuras y vehículos conceptuales. En §2 sugeriré un modo de articulación de los distintos tipos de representaciones semánticas en el pluralismo amplio y señalaré las ventajas de esta propuesta en relación a cada uno de los requisitos fundamentales.

En §3 analizaré las relaciones entre tipos de vehículos conceptuales y tipos de procesamiento. Asimismo, reconstruiré las tesis centrales de las arquitecturas duales, que parecen adecuadas para explicar la interacción entre los distintos tipos de vehículos, y sugeriré adoptar la versión de Sloman (1996, 2002, 2014), según la cual cada uno de los sistemas se caracteriza por tener representaciones en un formato particular y un tipo característico de procesamiento. Mientras que uno es fundamentalmente asociativo y sus operaciones reflejan similitud y estructura espacio-temporal, el otro es simbólico y opera a partir de reglas formales sensibles a la estructura sintáctica. En §4 propondré cómo articular esta versión de arquitectura dual con las tesis principales del pluralismo amplio. Señalaré que la propuesta resultante es similar a la distinción que propone Carey (2004, 2009) entre cognición nuclear y cognición en general. Así, habría dos sistemas conceptuales: uno nuclear, conformado por representaciones de tipo perceptivo, que serían los *outputs* de sistemas analizadores innatos y otro, posterior en el desarrollo, formado por representaciones de tipo simbólico. En ambos sistemas podrían, a su vez, identificarse estructuras semánticas de distinto tipo. En §5 presentaré un caso de aplicación de una arquitectura similar para la adquisición de conceptos de números naturales y, a partir de este ejemplo, ofreceré una explicación de la adquisición de conceptos que tiene en cuenta los distintos tipos de vehículos y estructuras semánticas. Por último, en §6 destacaré las ventajas de la combinación del pluralismo amplio y una arquitectura dual, y anticiparé una serie de posibles objeciones a las que ofreceré respuestas.

1. La necesidad de postular un pluralismo amplio a partir de los requisitos explicativos fundamentales

En los capítulos 2 a 5 evalué los tipos principales de estructura conceptual propuestos en la literatura y sostuve que ninguno puede dar cuenta de los requisitos explicativos fundamentales, i.e. *intencionalidad*, *alcance*, *composicionalidad*, *categorización*, *adquisición* y *publicidad*. De hecho, cada enfoque suele priorizar algunos requisitos por sobre otros y, en consecuencia, concentrarse en explicar un conjunto específico de fenómenos postulando las entidades que más se adecúan a ellos. Al analizar algunos de los requisitos fundamentales establecidos en el capítulo 1, puede observarse que imponen ciertas restricciones sobre el tipo de representaciones o estructuras que permiten explicarlos. Esto no resulta extraño. Los requisitos corresponden a fenómenos cognitivos y dar cuenta de ellos implica explicar determinadas capacidades, responder a ciertos argumentos relativos fundamentalmente a la composicionalidad y al contenido, y acomodar los distintos efectos empíricos observados en

torno a las capacidades en cuestión. En consecuencia, no cualquier tipo de estructura postulada resultará adecuada.

Los fenómenos empíricos asociados a las tareas de *categorización*, como vimos en los capítulos 2 a 5, parecen requerir de cierta estructura semántica para su explicación, ya sea una estructura constitutiva de los conceptos o sólo asociada a ellos. Esto es así dado que para explicar este proceso es necesario apelar a una estructura de rasgos sobre la que pueda operar una función de similitud, tal como un conjunto de ejemplares o un prototipo, o a una estructura de relaciones causales, que permita explicar la categorización como una forma de razonamiento (como un razonamiento deductivo o una inferencia a la mejor explicación). Se podría intentar explicar este proceso como comparación entre representaciones atómicas. Pero de ese modo, no podrían explicarse, *inter alia*, las diferencias en los tiempos de reacción asociadas a la tipicidad de las instancias, que vimos en el capítulo 2, o las diferencias frente a escenarios de transformación entre clases naturales y artefactos, que vimos en el capítulo 3. En este sentido, aun si se quisiera postular que los conceptos son atómicos, y que todos estos efectos se explican por representaciones asociadas a ellos, pero no constitutivas, sería preciso explicar cómo intervienen estas representaciones asociadas en los procesos que conectan instancias en el mundo con representaciones atómicas en la mente.

Observaciones análogas podrían formularse respecto de la *adquisición* y el *cambio conceptual* a lo largo del desarrollo. La adquisición impone la incorporación de ciertos mecanismos de acceso semántico (basados en última instancia en mecanismos perceptivos), para poder explicar cómo nuestras mentes entran en contacto con las propiedades en el mundo. De modo que esto demanda una articulación entre los conceptos y los mecanismos perceptivos, que son los que posibilitan el acceso al mundo y así a los contenidos de los conceptos. El cambio conceptual en el desarrollo (ya sea que se considere que hay continuidad o discontinuidad conceptual entre los niños y los adultos) requiere que, por lo menos en algunos casos, la información acerca de una categoría pueda modificarse a lo largo del tiempo (total o parcialmente, dependiendo de la postura adoptada).

La *composicionalidad*, a su vez, impone una restricción sobre los vehículos, dado que para poder combinarse de modo productivo y sistemático éstos deben tener estructura sintáctica, i.e. deben ser simbólicos (Fodor 2008). El estatus de los argumentos a favor de la composicionalidad, puede parecer ambiguo, en tanto parece imponer restricciones *a priori* sobre las características conceptuales. Sin embargo, la productividad y sistematicidad no pretenden incorporar compromisos teóricos sin fundamento empírico, sino que generalmente se los considera fenómenos empíricos asociados con nuestras capacidades representacionales.

Aunque no se trata de fenómenos respecto de los cuales se lleven a cabo estudios empíricos, se los suele aceptar como características de nuestro pensamiento que hay que explicar y se argumenta, a partir de ellos, que el sistema conceptual debe poseer composicionalidad (más allá de las interpretaciones más fuertes o más débiles que puedan ofrecerse de esta afirmación). Es en este sentido que los distingo de los distintos efectos empíricos de los que las teorías de conceptos deben dar cuenta.

Por otra parte, los requisitos están relacionados, en el sentido de que las tesis que se asuman con respecto a algunos de ellos acarrearán consecuencias relativas al resto. Así, por ejemplo, el requisito de *publicidad*, impone ciertas restricciones a la explicación de la *intencionalidad*. La individuación del contenido de los conceptos debe darse de un modo tal que sean compartibles. Como vimos en los capítulos 2 y 3, algunas teorías semánticas pueden conducir al holismo, i.e. a un sistema en el que el contenido de un concepto depende de sus relaciones con todo el resto de los conceptos. Así, la adopción de ciertas tesis relativas al contenido podría impedir la explicación de la publicidad, lo cual, a su vez, implicaría que no se puedan formular generalizaciones psicológicas intencionales sobre distintos individuos o sobre un mismo individuo en distintos momentos y que no se pueda explicar la comunicación lingüística.

Así, los distintos requisitos explicativos comprenden una serie de fenómenos variados e interrelacionados a los que deben adecuarse las estructuras conceptuales postuladas, y esto restringe los tipos de representaciones que resultarán adecuadas. Dadas estas consideraciones, los requisitos mismos parecen apuntar a distintos tipos de representaciones, de modo que parecen sugerir un pluralismo. Esto, por supuesto, adoptando el conjunto de requisitos propuesto, el cual no está exento de debate. Muchos de los enfoques priorizan algunos requisitos y sacrifican otros. Sin embargo, como sostuve en el capítulo 1, aunque un tipo de estructura particular se proponga para explicar, por ejemplo, la categorización, es razonable preguntarse cómo esa estructura adquiere contenido, si resulta adecuada para todos los tipos de conceptos que podemos tener, si posee composicionalidad, en síntesis, si satisface el resto de los requisitos explicativos.

De la misma manera que la adopción de un conjunto de requisitos explicativos requiere fundamentación, la intención de ignorar algunos de ellos generalmente aceptados en la literatura también. Así, por ejemplo, Fodor, como parte de la defensa del atomismo informacional, pretende demarcar las áreas de “injerencia” de las teorías metafísicas de conceptos y las teorías psicológicas de conceptos y sostiene que ha sido un error de la ciencia cognitiva el haber prestado excesiva atención a la categorización (Fodor 1999, 2000b). Fodor

sostiene esto debido a que las teorías que se han propuesto para explicar los efectos asociados a la categorización postulan conceptos internamente estructurados, lo cual, según Fodor, implica un compromiso con una semántica de rol conceptual que, a su vez, acarrea consecuencias holistas.

El hecho de que la psicología cognitiva haya investigado los conceptos principalmente *a través* de la categorización puede haber provocado ciertas distorsiones en el estudio de los conceptos (Keil & Wilson 2000). Esto es algo que los psicólogos cognitivos están dispuestos a admitir. Sin embargo, lo que conduce a distorsiones es, en todo caso, el hecho de ignorar fenómenos como el contenido y la composicionalidad, no el hecho en sí de explicar la categorización. El hecho de que la atención a la categorización sugiera que los conceptos están internamente estructurados, no es razón suficiente para rechazar por completo los fenómenos asociados a la categorización como parte de lo que una teoría de conceptos deba explicar. En todo caso, la objeción debería enfatizar la necesidad de incorporar consideraciones relativas al contenido, la composicionalidad y el razonamiento. Este es el camino que sugieren adoptar, por ejemplo, Keil y Wilson (2000).

Así, en los capítulos 2 a 5 evalué los principales enfoques respecto de la estructura y los vehículos conceptuales en relación a cada uno de los requisitos fundamentales (y no sólo en términos de los requisitos que cada enfoque prioriza). No sólo el análisis *a priori* de los requisitos, sino también la evidencia empírica de distintas investigaciones, parece apuntar en la dirección del pluralismo. Ahora bien, éste podría ser un pluralismo de estructura interna o de vehículos. Esto es, la propuesta de que el contenido de los conceptos puede estar estructurado de distintas maneras, por ejemplo según rasgos definicionales o estadísticos, y en consecuencia serán respectivamente definiciones o, por ejemplo, prototipos, o un pluralismo de vehículos, esto es, la propuesta de que el contenido puede estar vehiculado por distintos tipos de representaciones, en particular por representaciones perceptivas y representaciones simbólicas. Estas dos tesis son independientes, de manera que adoptar una no implica un compromiso con la otra. Se podría adoptar un pluralismo de estructuras semánticas y una uniformidad de vehículos o un pluralismo de vehículos que tengan la misma estructura semántica interna. Denomino a mi propuesta *pluralismo amplio*, porque abarca ambos tipos de pluralismo, a favor de los cuales considero que hay tanto argumentos como evidencia empírica.

Respecto al pluralismo semántico, se puede elaborar un argumento a su favor a partir de una inferencia a la mejor explicación que toma en cuenta tanto los distintos efectos observados en las investigaciones en torno a los procesos de categorización, de adquisición y

de cambio conceptual, como ciertas consideraciones relativas al contenido y a la composicionalidad. Con respecto a los primeros, los resultados de los distintos experimentos en torno a las tareas de categorización, discutidos en el capítulo 2, no pueden ser acomodados por un único tipo de estructura. Así, no hay un único tipo de estructura que sea la mejor disponible para dar cuenta de todos los fenómenos que se observan en estas tareas. De modo que se puede sostener que la mejor explicación de estos fenómenos tomados en conjunto consiste en postular distintos tipos de estructuras semánticas (Weiskopf 2009a).

Las teorías de prototipos y ejemplares no pueden acomodar los efectos causales, mientras que la teoría-teoría no puede acomodar los efectos de tipicidad y de familiaridad con instancias. La teoría de prototipos no puede acomodar los efectos de atipicidad, y la teoría de ejemplares debe apelar a una representación de tipo prototípico para acomodar los efectos de tipicidad (la cual se puede incorporar al enfoque de ejemplares por un mecanismo de abstracción de la tendencia central del conjunto de ejemplares almacenados que genere así una representación prototípica). Una vez aceptado esto, hay dos actitudes posibles que se pueden adoptar. Se puede seguir buscando una estructura alternativa que consiga acomodar todos estos fenómenos, o asumir que por lo menos en el caso de la categorización, esta tarea se resuelve haciendo uso de distintos tipos de estructuras. Esto último sería equivalente a afirmar que la mejor explicación de los fenómenos observados en la categorización requiere postular un pluralismo de estructuras semánticas. Se pueden formular argumentos análogos para el resto de los requisitos.

Respecto del pluralismo de vehículos, hay múltiples experimentos que intentan mostrar que los vehículos del pensamiento son perceptivos, pero también hay poderosos argumentos a favor de los vehículos simbólicos. Como vimos en el capítulo 5, a propósito del neo-empirismo de conceptos, en las últimas décadas se han obtenido resultados en múltiples estudios con distintos tipos de actividades, que sugieren que los procesos perceptivos y los procesos conceptuales comparten mecanismos, i.e. que llevar a cabo operaciones conceptuales implica activar sistemas perceptivos. Distintos autores interpretan estos resultados como evidencia a favor de la tesis de que los vehículos conceptuales son representaciones perceptivas (Barsalou 1999, Barsalou *et al.* 2003, Pecher 2013, Prinz, 2002). Ahora bien, los vehículos perceptivos enfrentan principalmente dos líneas de objeción. Por un lado, como vimos en el capítulo 5 en relación a las dificultades de la teoría de proxitipos con el requisito de alcance, la base perceptiva no consigue acomodar conceptos abstractos y teóricos, sino que resulta adecuada únicamente para conceptos de objetos perceptibles. Por otra parte, y como Barsalou y sus colegas (Barsalou 1999, Pecher 2013, Prinz 2002) de hecho

admiten, un modelo amodal de representaciones puede acomodar los resultados citados si postula, por ejemplo, que la información conceptual está organizada espacialmente en el cerebro por modalidades. Sin embargo, como argumenté en el capítulo 5, no toda la evidencia resulta fácilmente acomodable por una hipótesis de este tipo. A continuación presentaré esta crítica en mayor detalle y ofreceré una respuesta a la misma.

Machery (2006, 2007) cuestiona la posibilidad de extraer conclusiones relativas a los vehículos modales o amodales en general a partir de estudios conductuales, dado que este tipo de estudios sólo sirven para descartar modelos específicos diseñados para tareas específicas. Machery señala que, puesto que distintos modelos amodales de procesamiento arrojan distintas predicciones para una determinada tarea, los resultados que supuestamente favorecerían los vehículos modales deberían mostrarse inconsistentes con *todas* las predicciones de los distintos modelos amodales, lo cual no es posible. El problema que revela Machery es el de la puesta a prueba de enfoques o tesis generales. Sin embargo, el problema sería utilizar la evidencia a favor de los vehículos modales para descartar las representaciones amodales, i.e. utilizarla como evidencia en contra del enfoque amodal en general.

Lo que propongo, en cambio, es utilizar esta evidencia para afirmar que ciertos fenómenos relacionados con ciertas tareas se explican mejor apelando a representaciones perceptivas que a representaciones amodales. Fundamentalmente los efectos de interferencia y *priming* entre procesos perceptivos como la detección perceptiva o el recuerdo de cierta información perceptiva y procesos conceptuales como la verificación de propiedades, se explican mejor apelando a los mismos mecanismos subyacentes a una y otra tarea que a dos mecanismos distintos que operan sobre representaciones con distintos formatos. Resulta consistente con un enfoque pluralista que un tipo de vehículo sea el más adecuado para resolver ciertas tareas, aunque no necesariamente para todas. Podría ser que, entre otras tareas, para verificar propiedades perceptivas de los conceptos recurramos a representaciones perceptivas de modalidad específica, como sugieren los estudios revisados en el capítulo 5.

Las razones para postular vehículos amodales, se basan en los argumentos que vimos en el capítulo 4 en torno al tipo de vehículo que posee composicionalidad sintáctica y que permite realizar razonamientos de tipo formal (Fodor 2008), así como en las dificultades que tiene la teoría de proxitipos, desarrolladas en el capítulo 5, para dar cuenta de la individuación del contenido y el alcance sin recurrir a representaciones amodales. En este sentido, pueden señalarse las ventajas que ofrecen los vehículos amodales para la representación de conceptos abstractos y teóricos, y para dar cuenta de los procesos de razonamiento dependientes de la forma.

A la luz de estas consideraciones, creo que una teoría que adhiriera tanto al pluralismo de estructuras semánticas, como al pluralismo de vehículos presentará considerables ventajas explicativas respecto de aquellas que postulan un único tipo de estructura semántica o un único tipo de vehículo. Así, el *pluralismo amplio* afirma que los conceptos incluyen definiciones, prototipos, ejemplares, teorías y etiquetas atómicas y, por otro lado, propone que los conceptos pueden estar vehiculados tanto en un formato simbólico como perceptivo.⁵² Para explicar los fenómenos de intencionalidad y composicionalidad, los conceptos de tipo simbólico y sin estructura interna resultan ser los más adecuados. Sin embargo, para explicar los fenómenos relacionados con los procesos conceptuales, i.e. la categorización y la adquisición, los conceptos estructurados resultan más explicativos. Respecto de la adquisición, los vehículos perceptivos permiten establecer un vínculo entre la percepción y la cognición, de modo que resulta ventajoso incluir este tipo de vehículos en la teoría. Y en relación a la categorización, lo importante para poder dar cuenta de ella es la estructura de rasgos de los conceptos, cualquiera sea el vehículo que se proponga para transportarla.

Una propuesta de pluralismo amplio resulta atractiva porque promete heredar las ventajas explicativas de los distintos tipos de estructuras y vehículos que incorpora. Sin embargo, de la misma manera que heredaría sus ventajas, presenta el riesgo de heredar todas sus dificultades. En lo que sigue señalaré las ventajas de esta propuesta en relación a cada uno de los requisitos y sugeriré cómo podría hacer frente a las dificultades, en especial, a las críticas que formulé en los capítulos anteriores respecto de las propuestas que adhieren a la tesis de la uniformidad.

2. El poder explicativo del pluralismo amplio

En esta sección intentaré mostrar las ventajas explicativas del pluralismo amplio en relación a cada uno de los requisitos fundamentales. Este no pretende constituir un desarrollo exhaustivo y acabado que explique cómo dar cuenta de cada uno de estos requisitos, sino una sugerencia especulativa de cómo sería posible satisfacerlos desde la propuesta del pluralismo amplio.

⁵² Esta tesis es una propuesta general que postula distintos tipos de estructuras y pretende retomar las características principales atribuidas a cada una de ellas por los defensores de los distintos enfoques. Un trabajo de investigación a futuro consiste en la elaboración de modelos específicos para cada tipo de estructura postulada, así como de la interacción entre las distintas estructuras. En especial en relación a la teoría-teoría, es preciso especificar si se postula una versión de los conceptos en teorías o de los conceptos como mini-teorías.

Categorización

El requisito de categorización es el que ofrece motivos más claros para adoptar un pluralismo: el requisito de categorización. Los estudios empíricos citados en los capítulos 2 y 3 sugieren que no hay un único tipo de estructura que pueda acomodar todos los efectos que se observan en las tareas de categorización, de modo que, para resolver esto, se postulan distintas estructuras. Sin embargo, ¿cómo se explica qué estructura se utiliza en cada caso? Constituye un desafío de las propuestas pluralistas (y eliminativistas) explicar qué es lo que determina la activación de cada tipo de estructura. Un buen indicio de la que podría ser la clave para responder esta pregunta está en los mismos estudios citados. Las tareas que se proponen a los participantes en los distintos estudios no son las mismas y no involucran en todos los casos el mismo tipo de categoría. Presumiblemente, las tareas que involucran casos actuales no sean de la misma naturaleza que las que involucran casos hipotéticos. Y, presumiblemente, la categorización de primer orden o de categorías básicas, como MANZANA o ROJO, sea distinta a la de segundo orden, como en el caso de FRUTA o COLOR. Lo que intento sugerir es que una respuesta posible a qué tipo de estructura se activa para categorizar podría depender del orden o nivel de la categoría en cuestión y de las condiciones específicas de cada tarea. Gran parte de la investigación en categorización trabaja con categorías naturales. Sin embargo, no siempre se evalúa la estructura de las representaciones de categorías del mismo nivel y el tipo de tarea presentada varía.

En los estudios de Rosch y Mervis (1975), por ejemplo, las tareas de categorización consistían en determinar si instancias de una categoría de nivel básico correspondían a una categoría supraordinada, i.e. una categoría más inclusiva, con mayor grado de abstracción (por ejemplo, indicar si las sillas, mesas y lámparas caen bajo MUEBLE). En cambio, los estudios que proponen escenarios de transformación, como los llevados a cabo por Keil (1989), en los que se proponía que un grupo de doctores había realizado ciertas operaciones para cambiar la apariencia de distintos elementos (por ejemplo, habían modificado un caballo para que pareciera una cebra), trabajan con la atribución de una categoría de nivel básico a una instancia en un escenario contrafáctico.

Otros estudios trabajan con categorías artificiales. Tal es el caso de la mayoría de los llevados a cabo por los defensores del enfoque de ejemplares, así como algunos del enfoque de prototipos. En particular, los realizados por autores que se proponen evaluar predicciones alternativas de los enfoques de prototipos y ejemplares, en los que se ofrecen como estímulos instancias de categorías artificiales, como conjuntos de puntos dispuestos en algún patrón (Smith J. D. 2002), series de letras sin significado (Smith & Minda 1998) o combinaciones de

formas y colores (Medin & Schaffer 1978), de modo que los individuos deben primero adquirir representaciones de una categoría nueva y luego, en una segunda fase, categorizar instancias novedosas.

Resulta razonable que ante la presentación de conjuntos de puntos o formas y colores, las categorizaciones se guíen por las características superficiales, mientras que en los escenarios de transformación entren en juego las relaciones causales entre rasgos, así como la idea de una esencia oculta para las categorías biológicas. Pero esto no significa que, en todos los casos de categorización de instancias de cebras o caballos recurramos a nociones teóricas. Puede sostenerse que el tipo de proceso que llevamos a cabo para categorizar, y el tipo de información que activamos para la tarea, está influido por el contexto, por lo menos, en dos sentidos. Primero, la categorización involucra distintos tipos de tareas. En términos generales, involucra la atribución de una categoría a una instancia y la anticipación de las propiedades que tendrá una instancia por pertenecer a una categoría. Estas dos tareas no necesariamente se llevan a cabo mediante los mismos procesos, de modo que introducen una primera dimensión de variabilidad en las estructuras que se activan para llevarlas a cabo. Además, cada una de estas tareas puede variar en múltiples dimensiones. Por ejemplo, se puede manipular el tipo de respuesta a la tarea (i.e. podría ser verbal, o un movimiento).

Segundo, los estímulos o instancias que se pretende categorizar (o cuyas propiedades se pretende anticipar), también pueden ser altamente variables. No es lo mismo percibir, e.g. una foto de un caballo y tener que asignarle una categoría, que escuchar un relato sobre las transformaciones efectuadas en un caballo y luego tener que determinar a qué categoría pertenece. Probablemente la respuesta no sólo dependa de la información almacenada acerca de los caballos, de cómo está estructurada y de qué tipo de procesos operan sobre ella, sino también de la información presentada en los estímulos específicos de la tarea. Es posible que frente a información de rasgos superficiales, y sólo de rasgos superficiales, no necesitemos activar una red compleja de relaciones causales, sino que sea suficiente una comparación de tales rasgos con una representación prototípica. Esto no va en contra de los casos en los que, a pesar de las transformaciones superficiales, se juzga que una instancia de una categoría biológica, no cambia de categoría. No contradice tales casos, porque en esos casos, el escenario empieza afirmando que algo es un caballo, y esta categoría inicial no está puesta en duda. En este sentido, el escenario de transformación puede inducir la activación de información de distinto tipo que la categorización a partir de la percepción visual.

La sugerencia es que no ponemos en juego todo el conocimiento de teorías biológicas cuando nos cruzamos, por ejemplo, con un gato. No nos preguntamos si tiene esencia de gato,

o si no será una liebre “camuflada” para un experimento particular. Parece adecuado afirmar que en nuestra vida cotidiana nos guiamos por las características superficiales de las instancias en nuestras interacciones con ellas. Puede señalarse que la mayoría de los estudios de la teoría-teoría involucran tareas en las que la información presentada, o bien ya incluye relaciones causales explícitas (por ejemplo los estudios de Ahn y sus colegas de las hormigas *kehoe*, visto en el capítulo 3), las cuales pueden conducir a estructurar la información de una categoría a partir de estas relaciones, o implica escenarios hipotéticos, que pueden por sí mismos, más allá de cómo esté estructurada la información acerca de las categorías, inducir que se lleven a cabo razonamientos. En este sentido, qué tipo de información se usa (un prototipo, un conjunto de ejemplares o un conjunto de relaciones causales o principios teóricos), y qué tipo de proceso se lleva a cabo (una comparación de rasgos o una inferencia a la mejor explicación) depende, por un lado, del tipo de tarea que se lleva a cabo y, por otro lado, de la información recibida acerca de la instancia que se quiere categorizar.

Ahora bien, ¿qué ocurre si la categorización mediada por un prototipo y aquella mediada por una teoría conducen a resultados distintos? Recordemos que la posibilidad de que se produzcan estos casos era uno de los argumentos de Machery a favor de las teorías pluralistas y en contra de las híbridas, discutido en el capítulo 6. Si bien, como sostuve, no coincido con Machery en que este tipo de resultado juegue en contra de las teorías híbridas, acuerdo en que esto es algo de lo que hay que dar cuenta. De hecho, es prácticamente el punto de partida de una propuesta pluralista que las representaciones definicionales, prototípicas, de ejemplares, teóricas y atómicas son representaciones distintas que, si bien refieren a la misma categoría, también tienen una estructura interna que las distingue unas de otras. Y esto puede conducir a que, frente a una misma instancia, se den casos de categorizaciones cruzadas, en las que, por ejemplo, de acuerdo a un prototipo la instancia pertenece a una categoría y de acuerdo a una teoría, a otra.⁵³

Para abordar esta cuestión, resulta útil una distinción que ofrece Machery (2009) entre competencia cognitiva, proceso cognitivo y tarea. Machery caracteriza las competencias cognitivas en términos de funciones. Por ejemplo, el reconocimiento visual de rostros es una capacidad cognitiva, así como la distinción de fonemas o la estimación de la cardinalidad de un conjunto de objetos. Las capacidades cognitivas suelen involucrar subcompetencias. Por ejemplo, la distinción de objetos tridimensionales en el campo visual involucra la capacidad

⁵³ Esta tesis acarrea ciertas dificultades relativas al contenido, puesto que habría representaciones supuestamente correferenciales que, sin embargo, parecen recortar el mundo de distintas maneras. Me ocuparé de esta cuestión al tratar el requisito de contenido.

de identificar las sombras que proyectan estos objetos (Machery 2009). Un proceso cognitivo, a su vez, es un modo de llevar a cabo la función de una determinada competencia cognitiva y, en este sentido, los procesos cognitivos subyacen a las competencias cognitivas. La relación entre capacidades y procesos no necesariamente es una relación uno a uno. Podría haber distintos procesos cognitivos subyacentes a una determinada competencia. Machery explica esto por analogía con las habilidades. Por ejemplo, manejar un auto es una habilidad. Pero hay distintos modos de llevarla a cabo, e.g. se puede manejar por la derecha o por la izquierda, dependiendo del lugar del mundo. Al ejecutar la habilidad de un modo u otro, los procesos involucrados se modifican, e.g. el cálculo del tamaño del auto desde la posición del conductor es diferente, la palanca de cambios se maneja con la mano izquierda o con la derecha, etc.

Por último, las tareas son “situaciones experimentales en las que se espera que los sujetos lleven a cabo alguna acción” (Machery 2009, p. 122). Hay un sentido intuitivo en el que no hay dudas respecto de qué es una tarea y cómo se distingue una tarea de otras. Indicar si una determinada instancia corresponde a una categoría o no, responder si las instancias de una categoría tienen una determinada propiedad son ejemplos de tareas. La cuestión de interés es determinar en qué casos se trata de la misma tarea. Por ejemplo, responder si una oración tiene sentido levantando la mano o expresando “sí” o “no” en voz alta o presionando un botón, ¿corresponden a la misma tarea? Como ilustran los estudios analizados en los capítulos 2 a 5, en ciencia cognitiva se proponen distintos tipos de tareas en contextos experimentales con el objetivo de poner a prueba determinadas hipótesis acerca de los procesos que subyacen a las competencias cognitivas que son objeto de estudio. Se supone que para llevar a cabo la tarea propuesta, se ejecutarán determinados procesos y que ciertos parámetros observables en el estudio permitirán extraer conclusiones u obtener indicios acerca de las características de tales procesos.

Así como la relación entre competencias y procesos no parece ser uno a uno, la relación entre procesos y tareas tampoco. Una misma tarea puede ejecutarse por medio de procesos distintos y el mismo proceso puede usarse también para distintas tareas. Gran parte del esfuerzo puesto en el diseño de un determinado estudio se relaciona con delimitar la tarea lo máximo posible, de modo de inducir que se lleve a cabo por medio del proceso específico que se pretende estudiar. Un ejemplo claro de esto está dado en los estudios relativos a los vehículos que se proponen evaluar la interacción entre los sistemas perceptivos y los conceptuales, generando una interferencia al imponer una carga de representaciones perceptivas en la memoria, en simultáneo con una tarea conceptual. El diseño de esta tarea

particular pretende que se activen determinadas representaciones perceptivas, de modalidad específica, para evaluar si el proceso de verificación de propiedades se ve afectado por tal activación. El hecho de que se observe interferencia, sugiere que la carga en la memoria ocupa recursos que el proceso conceptual requiere.

De modo que cómo se individualúan las competencias, los procesos y las tareas tiene consecuencias respecto de las conclusiones que se puedan obtener por medio de los estudios empíricos. Si una determinada competencia es llevada a cabo por procesos radicalmente distintos, que dan lugar a resultados distintos, ¿en qué sentido son procesos que subyacen a una misma competencia? o ¿se trata de competencias distintas? Las mismas preguntas se pueden formular para procesos y tareas. En particular, en relación al requisito de categorización y retomando la sugerencia de que distintos tipos de tareas conducen a la activación de distintos tipos de representaciones y procesos, es preciso ofrecer criterios de individuación de competencias, procesos y tareas. La categorización de un conjunto de puntos a partir de sus rasgos perceptivos y la categorización de un individuo que era un caballo y atravesó un proceso complejo de operaciones de transformación, ¿involucran la misma competencia cognitiva o dos competencias distintas? Lo mismo puede preguntarse respecto de los procesos. Podría afirmarse que el proceso de comparación de una instancia con un prototipo o un conjunto de ejemplares o la explicación de sus atributos a partir de una red causal constituyen procesos distintos o, alternativamente, que son variantes de un mismo tipo de proceso.

Machery sugiere cómo individualuar las capacidades, los procesos y las tareas. Respecto de las competencias cognitivas, sugiere que si al ser capaz de llevar a cabo una determinada función en un contexto se es capaz también de llevarla a cabo en otro, entonces la ejecución de la función en los dos contextos involucra la misma competencia. En cambio, si al ser capaz de llevar a cabo una función en un contexto no se es capaz de llevarla a cabo en otro contexto, se trata de dos competencias diferentes. Machery sostiene que este último sería el caso si, para llevar a cabo la función en un contexto diferente, se deben inhibir los mecanismos que se habían automatizado para ejecutar la función en el primer contexto. En este caso, se trata de una competencia nueva y no de la misma en un contexto distinto, puesto que ejecutarla implica modificaciones en los mecanismos que se habían automatizado.

Con respecto a los procesos cognitivos, Machery recurre a tres criterios propuestos por Ashby y Ell (2002). Ashby y Ell distinguen sistemas apelando a criterios de tres niveles: matemático, psicológico y neurobiológico. En el nivel matemático, dos modelos de dos sistemas son distintos si las predicciones de ambos no son equivalentes y ninguna es un

subconjunto de la otra, i.e. si realizan por lo menos alguna predicción única. En este nivel, los sistemas se individualizarían como funciones de *input-output*.⁵⁴ En el nivel psicológico, dos sistemas son distintos si postulan que se requieren distintos procesos para llevar a cabo una tarea dada de manera exitosa, i.e. si proponen distintos mecanismos o algoritmos específicos para ejecutar una función determinada. Y, por último, en el nivel neurobiológico, dos sistemas son distintos si están implementados en estructuras o vías neurales independientes. Ashby y Ell sugieren que en un extremo estarían los sistemas que satisfacen los tres criterios, los cuales serían sin lugar a dudas sistemas distintos, en el otro extremo, aquellos que no satisfacen ninguno, que serían sistemas idénticos y, en el medio, un continuo de sistemas que satisfacen alguno o algunos de estos criterios, pero no los tres.

Machery propone adoptar estos criterios y sugiere, para el caso específico de la categorización, que los procesos de categorización vía prototipos, ejemplares o teorías, contarían como procesos distintos si se cumplieran estos tres criterios o algunos de ellos. Con respecto a las funciones *input-output*, para algunos casos específicos, los modelos propuestos por los enfoques de prototipos, ejemplares y teoría-teoría predicen respuestas distintas. Sirven de ejemplo los casos de los que se ocupan Machery y Seppälä (2011), discutidos en el capítulo 6. Así, tal como se suelen modelar los procesos que operan sobre prototipos, ejemplares y teorías, pueden encontrarse casos para los cuales estos modelos arrojan predicciones distintas. Con respecto al criterio psicológico, como vimos, los tipos de operaciones que se proponen para categorizar mediante prototipos, ejemplares y teorías son distintos. Para los primeros se sugiere que son operaciones de comparación de similitudes entre rasgos superficiales, mientras que para las teorías se proponen construcciones de explicaciones, i.e. se atribuye a una instancia la categoría que mejor explica sus rasgos. Así, para llevar a cabo la función de categorización, cada una de estas estructuras conlleva un algoritmo o procedimiento de tipo diferente y, en este sentido, siguiendo los criterios de Ashby y Ell, corresponderían a sistemas distintos.

Por último, con respecto a las estructuras o vías neurales que implementan los procesos que operan sobre cada tipo de estructura semántica para categorizar, la investigación aun no resulta suficiente para zanjar esta cuestión. Para afirmar que efectivamente estas estructuras y los procesos que operan sobre ellas están implementados en vías neurales distintas, se debería mostrar evidencia de dobles disociaciones entre sistemas neurales. Se debería encontrar evidencia de que un tipo de proceso puede operar independientemente del

⁵⁴ El criterio matemático, puesto que se basa en diferencias en las predicciones, permite distinguir entre modelos de sistemas y no estrictamente entre sistemas.

otro. En neuropsicología, este tipo de evidencia suele obtenerse en casos en los que pacientes con lesiones cerebrales en determinadas áreas exhiben déficits en un tipo de tarea o proceso, mientras que otros procesos se hallan intactos, y otros pacientes con otros tipos de lesiones cerebrales exhiben el patrón inverso. Si bien se han reportado casos de dobles disociaciones relativas a la implementación de las estructuras conceptuales, estos no permiten distinguir entre los procesos que operan sobre prototipos, ejemplares y teorías.⁵⁵ Así, en ausencia de evidencia neuropsicológica decisiva, las razones más fuertes para sostener que hay distintos tipos de procesos que subyacen a la categorización corresponden a los criterios de la modelización matemática y al psicológico.

Ahora bien, si como sugieren los modelos de definiciones, prototipos, ejemplares y teorías, los procesos que operan sobre cada una de estas representaciones son distintos, ¿qué determina que se active uno u otro? Una posibilidad es, como sugerí, que esto dependa de la tarea y el estímulo en cuestión. Esta opción requeriría una individuación de tipos de tareas. Como señalé, hay un sentido intuitivo en el que las diferencias en los tipos de actividad y estímulo propuestos involucran distintas tareas. Parece evidente que determinar si una manzana es una fruta y determinar si una hormiga que tiene sangre con un alto nivel de sulfato de hierro pero no digiere rápido es una hormiga Kehoe, son tareas distintas. Si bien ambas involucran nuestra habilidad para categorizar, no parece tratarse del mismo tipo de tarea.

En otros casos, la diferencia no resulta tan clara. Volvamos sobre algunos ejemplos vistos en el capítulo 5. No está claro si, por ejemplo, la tarea de responder si una oración tiene sentido pronunciado “sí” o “no” y la tarea de responder si una oración tiene sentido realizando un movimiento con la mano que se aleja del cuerpo para responder que sí y uno que se acerca al cuerpo para decir que no, son variaciones de la misma tarea. Parece innegable que el tipo de actividad que tiene que llevar a cabo el sujeto para responder a ambas tareas varía pero, a la vez, parece que la tarea de evaluación de sentido de la oración involucrada en ambas es la misma y, por ende, ambas tareas activarían el mismo proceso. Una distinción de tareas por consignas podría resultar demasiado fina, como ocurre con el ejemplo anterior. Como señala Machery (2009), no parece haber un modo no trivial de distinguir entre tareas. Individuar tareas resulta relevante porque las conclusiones que se pueden extraer de los estudios que las involucran implican supuestos respecto de los procesos que se activan para llevarlas a cabo. Parece esclarecedor afirmar que dos tareas son distintas

⁵⁵ Para una revisión de los principales estudios neuropsicológicos relativos a la representación de categorías y los procesos de categorización, véase Hanson y Hanson (2005).

si se llevan a cabo por medio de procesos distintos. Pero, puesto que la individuación de procesos, en la práctica, no puede darse independientemente de las tareas, este criterio resultaría circular. Dado que no se han formulado criterios explícitos de individuación de tareas, apelaré a una noción intuitiva, por lo menos para los casos más claros.

Retomando la cuestión de la activación de uno u otro proceso de categorización, una respuesta posible, como sugerí, apela a las tareas y los estímulos en cuestión. Machery (2009) señala otras posibilidades: que todos los procesos operen en paralelo y, o bien los distintos resultados se integren en una única respuesta, o bien se seleccione el resultado de uno de ellos por medio de algún mecanismo. Abordaré estas opciones en §4, al presentar un tipo de arquitectura cognitiva que resulta compatible con el pluralismo amplio y que cuenta con apoyo empírico independiente.

Contenido

El contenido es, probablemente, el requisito que resulta más desafiante para una propuesta de tipo pluralista. En primer lugar, un enfoque de este tipo obliga a desprenderse de la idea tradicional de que tenemos “el concepto PERRO” o “el concepto CASA” y a postular en cambio una multiplicidad de conceptos PERRO o CASA. Sin embargo, a la hora de explicar el contenido de estos conceptos, es relevante el hecho de que, por ejemplo, el prototipo $PERRO_P$, la teoría $PERRO_T$ y el símbolo atómico PERRO, son todos acerca de perros. De este modo, todas estas representaciones son acerca de la misma categoría. Pero, a la vez, están estructuradas a partir de información de distinto tipo acerca de esa categoría, de manera que respecto del contenido cognitivo, estas estructuras no serían equivalentes.

A lo largo de la tesis me referí a los prototipos, ejemplares y teorías como distintos tipos de estructura semántica interna, asumiendo que la estructura interna refleja un tipo de contenido o un aspecto del contenido de los conceptos. La estructura interna refleja contenido cognitivo, que constituye el modo en el que representamos una determinada categoría, i.e. recoge cómo está estructurada la información acerca de una categoría. La estructura interna es el modo en que están representadas las categorías, y, en este sentido, involucra cuestiones semánticas, pero a nivel interno, i.e., respecto de cómo está estructurada la información “en la cabeza”. Este tipo de consideraciones no implican, necesariamente, un compromiso con una teoría semántica específica respecto de cómo se relacionan los conceptos con el mundo. Sin embargo, puesto que tanto los enfoques de prototipos, como de ejemplares y de teorías asumen que los conceptos son representaciones de categorías del mundo, acarrearán compromisos semánticos también en sentido externalista, aunque sea de modo implícito.

Podría decirse, entonces, que a no ser que se adopte una semántica puramente internalista, la postulación de una estructura semántica interna trae aparejado un compromiso con una teoría semántica de doble factor. Según este tipo de semántica, hay dos factores que determinan el contenido de los conceptos: su relación con el mundo, qué categoría representan, y su estructura interna, i.e. cómo representan esa categoría.

De este modo, el problema que una propuesta pluralista debe enfrentar consiste en que los distintos tipos de estructuras semánticas “traen consigo” distintos modos de referir. Un símbolo atómico refiere en virtud de leyes causales que lo conectan con las instancias de la propiedad que representa. Una definición, en cambio, refiere por reglas, i.e. refiere a aquellas instancias que posean un conjunto de condiciones necesarias y suficientes para pertenecer a una categoría. Un prototipo, en cambio, refiere a sus instancias por similitud, dado que representa las características estadísticamente más frecuentes de la misma. Estas características no constituyen un conjunto de condiciones necesarias y suficientes para caer bajo la categoría, de modo que un prototipo no refiere únicamente a las instancias que tienen todas esas características, sino que refiere a todas las instancias que superen un *umbral* de similitud con ese conjunto de representaciones. En el caso de los ejemplares, cada uno refiere a una instancia particular de una determinada categoría. Esto puede explicarse apelando a una relación causal entre este ejemplar y la instancia que lo causó o a una relación de similitud entre ambos. En las teorías que postulan únicamente ejemplares, se supone que cada conjunto de ejemplares corresponde a una categoría en general en virtud de la similitud de todas las instancias de la categoría con uno u otro ejemplar almacenado. Las teorías, por otra parte, como vimos en el capítulo 3, no son suficientes por sí mismas para dar cuenta del contenido intencional. Si bien se puede sostener que disponemos de conocimiento acerca de relaciones causales y de esencias ocultas de las categorías, las teorías no siempre describen o especifican cuáles son esas esencias y la información causal que codifican no resulta suficiente para individuar categorías. Esto significa que, aunque las teorías codifican información acerca de la pertenencia a categorías, adoptar únicamente teorías como conceptos no permitiría dar cuenta del requisito de contenido.

Dada esta situación, ¿cómo articular estas distintas maneras de referir y resolver la dificultad de las teorías con este requisito? Un enfoque pluralista del tipo que defiende propone que hay, por un lado, conceptos amodales que refieren por relaciones causales, por otro lado, prototipos y ejemplares que refieren por similitud de rasgos y, por último, teorías que codifican información sobre las categorías pero que no refieren por sí mismas (sino en virtud de que incorporan otras representaciones que las conectan con el mundo), además de

definiciones para los conceptos matemáticos, que especifican un conjunto de condiciones necesarias y suficientes. Esto parece señalar que la mejor forma de articular estos distintos tipos de estructura es adoptando una teoría semántica de doble factor. Pero, como vimos en relación a la teoría de prototipos en el capítulo 5, esto no está exento de dificultades, de modo que es necesario ofrecer un enfoque de cómo se articulan todos estos tipos de contenido. Los principales problemas que surgen al postular dos tipos de contenido se relacionan con la individuación de los conceptos, dado que distintos conjuntos de representaciones pueden referir a la misma categoría y así compartir el contenido intencional, pero, a la vez, tener contenidos cognitivos distintos, por estar internamente estructuradas de diferentes maneras.

Una solución posible al problema de la individuación de conceptos se puede ilustrar mediante la metáfora de los archivos mentales, a la que me referí en los capítulos 4, 5 y 6 y que es cada vez más utilizada (Prinz 2002, Fodor 2008, Weiskopf 2009a). La idea de esta metáfora es que nuestro sistema conceptual estaría organizado en archivos, cada uno correspondiente a una categoría. Así, los distintos conceptos estarían agrupados o asociados en virtud de referir a la misma categoría (aunque esta referencia no sea del mismo tipo). Vimos que la semántica informacional permite dar cuenta del contenido intencional, de modo que la relación con el mundo estaría dada por medio de relaciones nomológicas con las etiquetas. Respecto del resto de las estructuras semánticas que permiten almacenar información sobre la pertenencia a categorías, considero que también puede adoptarse una semántica informacional para los ejemplares (dado que su manera de referir puede explicarse en virtud de relaciones causales con sus instancias), pero no para los prototipos, las teorías ni las definiciones. El contenido de estas representaciones no es producto de una mera relación causal, sino evidencial/racional, i.e. constituyen conocimiento acerca de las propiedades. En un sentido, este conocimiento es *acerca de* la propiedad de modo que también puede tomarse como refiriendo a propiedades, aunque no en virtud de relaciones meramente causales. A esto debe agregarse la salvedad señalada, de que es preciso o bien que haya representaciones conectadas con el mundo y asociadas a las teorías para que éstas tengan contenido intencional, o bien que las teorías mismas estén constituidas por conjuntos de rasgos causalmente conectados, y jerarquizados en función de rasgos centrales (que pueden ser esencias ocultas) y rasgos periféricos, constituidos por representaciones de otro tipo, ya sea atómicas, definiciones prototipos o ejemplares.

De modo que la propuesta que defiende consiste en individuar conceptos en virtud de su contenido cognitivo, i.e. por cómo representan aquello que representan. Así, por ejemplo, $PERRO_P$ y $PERRO_E$ serían conceptos distintos. Y se diferencian del símbolo atómico

PERRO_A, en tanto este último no tiene contenido cognitivo, sino únicamente intencional. En este sentido, esta propuesta reconoce la independencia de los distintos tipos de representaciones, que pueden activarse independientemente en la memoria de trabajo, y rechaza así la condición de asociación en sentido arquitectural, discutida en el capítulo anterior. A la vez, adopta una condición de asociación en sentido semántico, en tanto propone que los distintos conceptos correferenciales están asociados, puesto que forman parte del mismo archivo. Los archivos se individualarían en función de la propiedad que representan. En este sentido se puede agrupar PERRO_A, PERRO_P y PERRO_E, en virtud de que refieren a la propiedad *perritud*. De esta manera, los archivos serían conjuntos de conceptos que refieren a la misma propiedad.

Esta propuesta enfrenta, sin embargo, fundamentalmente dos dificultades. En primer lugar, es necesario ofrecer una explicación de cómo se determinan las asociaciones de los conceptos a cada archivo, lo cual no resulta sencillo. Y, en segundo lugar, si los archivos se individualaran sólo por su relación con el mundo no se podrían explicar los casos Frege. Esto resultaría paradójico, puesto que la incorporación de la información en los archivos, que sería contenido cognitivo, se hace en general precisamente para explicar este tipo de casos. Ambas dificultades están relacionadas. Este problema es el mismo que enfrentan las teorías de doble factor: cómo articular el contenido cognitivo y el intencional. Respecto a la primera cuestión, una respuesta posible es que las conexiones que asocian conceptos en un mismo archivo son de tipo epistémico, en tanto se basan en que tomamos a esos conceptos *como* refiriendo a la misma propiedad. Así, no sería el mundo el que determina causalmente esas asociaciones, sino nuestras creencias acerca de cómo es el mundo. De este modo, la asociación de conceptos a archivos estaría dada en función de cómo concebimos el mundo.

Esta respuesta puede contribuir a su vez a resolver la segunda dificultad, dado que la única manera de dar cuenta de los casos Frege es apelando a contenidos cognitivos distintos en archivos distintos. De este modo, podría haber dos archivos distintos que refieran a la misma propiedad en el mundo, pero serían archivos distintos porque contienen distintos contenidos cognitivos, i.e. contienen distintos ejemplares, teorías y prototipos. Por ejemplo, para los conceptos SUPERMAN y CLARK KENT, que refieren a la misma propiedad en el mundo, habría dos archivos con las etiquetas SUPERMAN_A y CLARK KENT_A respectivamente. Estas etiquetas están nomológicamente conectadas con la misma propiedad en el mundo. A su vez, cada archivo tendría un conjunto de conceptos distintos, porque tendría prototipos, ejemplares y teorías distintos que representan distintos aspectos de la categoría. Por ejemplo, el archivo SUPERMAN incluiría un ejemplar de un individuo con un

traje azul y los rasgos del prototipo de un superhéroe, mientras que el archivo CLARK KENT incluiría un ejemplar de un individuo con anteojos y los rasgos del prototipo de periodista. Por otra parte, habría que agregar asimismo que la apertura de uno u otro archivo en los primeros encuentros con la propiedad también debe depender de factores epistémicos, tales como qué mecanismos de acceso semántico intervienen en cada caso. De otro modo, esta propuesta tendría el mismo problema que el atomismo informacional, para el cual, como sostuve en el capítulo 4, no hay manera de explicar cómo una propiedad está nomológicamente relacionada con dos símbolos distintos.

Esta propuesta tiene puntos de contacto con los enfoques de Fodor (2008), Prinz (2002) y Weiskopf (2009b). Prinz propone una metáfora de archivos sin etiquetas, esto es, estructuras semánticas agrupadas pero sin símbolos amodales. Como vimos en el capítulo 5, muchas de las dificultades de la teoría de proxitipos se resolverían si incorporara representaciones amodales que sirvan de etiquetas para esos archivos. Mi principal crítica a la satisfacción del requisito de contenido para el neo-empirismo consistió en la necesidad de incorporar representaciones mentales amodales para poder dar cuenta de la individuación de conceptos. Lo que propongo es conservar la metáfora de los archivos, pero incorporando a los mismos las etiquetas amodales del atomismo informacional que permiten dar cuenta de casos en los que se tienen dos conceptos distintos con el mismo contenido cognitivo asociado (como el caso de dos individuos llamados Farrakhan que vimos en el capítulo 5). En estos casos, lo que permitiría diferenciarlos serían dos etiquetas amodales distintas, y la apertura de dos archivos distintos para la misma propiedad en el mundo (el mismo individuo), debería explicarse apelando a la intervención de distintos factores epistémicos en la apertura de uno y otro archivo.

La propuesta fodoriana, por otra parte, consiste en ‘asociar’ las etiquetas con archivos que almacenan información codificada en LDP. Esta propuesta no permite almacenar información prototípica, ejemplares, ni representaciones perceptivas, sino sólo conocimiento de tipo proposicional. Los archivos que propongo se diferencian de esta propuesta en tanto que incorporan otros tipos de estructuras y vehículos. Y, fundamentalmente, se diferencia del atomismo informacional, en que todas estas estructuras serían constitutivas de los conceptos y no meras ‘asociaciones’ a los mismos.

Además, como mencioné más arriba, esta propuesta retoma las dos tesis a las que, como sostuve en el capítulo 6, adhieren las teorías híbridas y pluralistas: (i) cada categoría puede estar representada por distintos conceptos correferenciales, que pueden activarse independientemente en la memoria de trabajo, y (ii) los distintos conceptos correferenciales

están asociados como representaciones de una misma categoría. Estas dos tesis requieren elucidación en relación al contenido. Como propongo, al igual que sugiere Prinz, cada conjunto de representaciones de un archivo que puede usarse para pensar acerca de la categoría que el archivo representa, cuenta como un concepto independiente. Esto es así porque puede activarse independientemente del resto y, en este sentido, es un concepto en sí mismo (en oposición a ser sólo una parte de él que depende del resto para funcionar). Los subconjuntos de un archivo tienen autonomía del resto en tanto se pueden usar independientemente. Esto recoge la tesis (i).

Ahora bien, la tesis de que distintos tipos de conceptos, que tienen una estructura semántica interna diferente y refieren de distinto modo, pueden ser correferenciales, se explica por las asociaciones epistémicas que asignan un conjunto de conceptos a un mismo archivo. Esto, a su vez, permite explicar por qué el caso $TOMATE_T$ - $TOMATE_P$ se diferencia del caso SUPERMAN-CLARK KENT. Sostuve que $TOMATE_T$ y $TOMATE_P$ son dos conceptos coextensivos en tanto representan la misma categoría, pero tienen distinto contenido cognitivo, dado que no sólo están estructurados internamente a partir de conjuntos de rasgos distintos, sino que también involucran distintos tipos de relaciones entre esos rasgos, y entre esos rasgos y el mundo. Ahora bien, la caracterización de dos conceptos como coextensivos y con distinto contenido cognitivo es el modo en que se describen los casos Frege. Sin embargo, no parece adecuado tratar el caso $TOMATE_T$ - $TOMATE_P$ del mismo modo que los casos SUPERMAN-CLARK KENT o YOCASTA-MAMÁ. Esto es así porque, según propongo, Edipo usa el prototipo y la teoría de tomate para pensar en la que para él es la misma categoría, mientras que usa YOCASTA y MAMÁ para pensar en las que para él son dos categorías distintas. De este modo, mi sugerencia es que se puede diferenciar entre casos Frege epistémicamente significativos y casos Frege espurios. Los conceptos que tienen distinta estructura interna, pero están asociados al mismo archivo, serían casos espurios, en tanto no traerían aparejadas consecuencias epistémicas significativas. Por esta razón resulta relevante la propuesta de incorporación de factores epistémicos en la individuación de archivos.

Los defensores de los vehículos modales, tal como vimos en el capítulo 5, recurren a argumentos de parsimonia para sostener que las teorías amodales no pueden prescindir de representaciones perceptivas, que son las que nos otorgan acceso semántico al mundo. En cambio, las teorías que adoptan vehículos modales no requieren de un nivel de representaciones en un formato distinto y pueden explicar todos los fenómenos conceptuales postulando únicamente representaciones perceptivas. Consideraciones análogas de

parsimonia también se han esgrimido en relación a la estructura conceptual para sostener que las representaciones atómicas resultan superfluas. Así, por ejemplo, Hampton, haciendo alusión a la adquisición de los símbolos atómicos y al rol de los prototipos en este proceso, señala:

La propuesta de que de algún modo se crean conceptos atómicos en nuestra mente mediante la exposición a instancias, identificadas como tales por otra parte de nuestra mente que ya ha creado prototipos para ellas, es realmente bizarra. ¿Por qué habría de haber evolucionado tal sistema representacional dual? Lo que es más importante, ¿no va a estar toda la acción interesante en el estudio de los prototipos, en lugar de en el de estas abstracciones no estructuradas? (2000, p. 300)

En respuesta a este tipo de argumentos, las consideraciones relativas al contenido refuerzan la idea de que las etiquetas cumplen una función individuadora. Si bien acuerdo con Hampton en que el rol de los prototipos en la adquisición requiere mayor atención y, como sostuve en el capítulo 4, la explicación que ofrece Fodor de la adquisición resulta insuficiente, esto no implica que las representaciones atómicas sean prescindibles. Hay fenómenos, en particular relativos al contenido, cuya mejor explicación involucra postular símbolos atómicos. Tal es el caso del ejemplo de FARRAKHAN₁ y FARRAKHAN₂, presentado en el capítulo 5, en el que un individuo cree que hay dos individuos llamados Farrakhan, sin tener ninguna representación específica asociada a ninguno de ellos que permita diferenciarlos. En estos casos, la postulación de una etiqueta para cada archivo ofrece un modo de distinguirlos. Esta etiqueta no necesita tener una estructura interna. De hecho, parece razonable que no podamos postular una estructura interna para todos los conceptos, sino que para algunas categorías no disponemos de estructuras de rasgos.

Composicionalidad

En relación a la composicionalidad, considero que se debe entender este requisito de manera debilitada, del modo que proponen Prinz (2002) y Robbins (2002). Así, lo que habría que mostrar es que los conceptos pueden combinarse de manera productiva y sistemática, pero no es necesario que siempre sea así. Respecto de este requisito, el pluralismo ofrece una clara ventaja. Las mayores críticas a la composicionalidad de los prototipos, ejemplares y teorías provienen del atomismo informacional. Pero dado que el pluralismo amplio incorpora

representaciones mentales de tipo simbólico, que pueden ser atómicas y análogas a las palabras de un lenguaje natural tal como postula Fodor, puede recurrir a ellas para hacer frente a estas críticas.

Por otra parte, el modelo RCA de Prinz (2002) es una buena sugerencia de cómo podrían contribuir distintos tipos de estructuras semánticas a la combinación de conceptos. En este modelo, como reconstruí en el capítulo 5, el uso de las distintas representaciones está organizado en etapas sucesivas que incorporan distintos mecanismos. Adoptar este modelo desde un enfoque pluralista permitiría evitar ciertas limitaciones que tiene el neo-empirismo. Estas limitaciones están vinculadas, por un lado, a la apelación a partes de los conceptos que no está claro que todos los conceptos tengan y, por otro lado, a la falta de representaciones atómicas que puedan garantizar la combinación para los casos de combinaciones novedosas en los que no disponemos de representaciones prototípicas, ejemplares previos almacenados en la memoria ni estructuras teóricas. Al adoptar un modelo similar pero desde un enfoque pluralista estos inconvenientes deberían disolverse, al mismo tiempo que se consigue aquello que Prinz sugiere con la lectura debilitada del requisito de composicionalidad: que la combinación estrictamente composicional sea un último recurso y no la estrategia general.

El hecho de postular representaciones atómicas garantiza la productividad y sistematicidad del pensamiento, puesto que éstas no están sujetas a las limitaciones en la combinación a las que están sujetos los prototipos, los ejemplares y las teorías. Así, habría distintas estrategias de combinación. A las tres etapas que propone el RCA se puede agregar una más, a la que recurriríamos en caso de que no tengamos representaciones semánticamente estructuradas para los conceptos que se quiere combinar. De este modo, para hacer frente a los famosos contraejemplos de Fodor a la composicionalidad de los prototipos, se puede recurrir a representaciones simbólicas atómicas. Este sería el caso, por ejemplo, para pensar ABUELAS LA MAYORÍA DE CUYOS NIETOS SON DENTISTAS. Podríamos disponer de una definición, un prototipo, un conjunto de ejemplares y una red causal para ABUELA, lo mismo para NIETO. Para DENTISTA, podríamos tener un prototipo, un conjunto de ejemplares y una red causal, y podría intentar explicarse la conformación del concepto compuesto como una combinación de los rasgos de los prototipos de ABUELA, NIETO y DENTISTA, más alguna modificación en la etapa de análisis a partir de una representación teórica del concepto de mayoría. Sin embargo, no parece que este concepto esté constituido por un conjunto de rasgos al que se pueda arribar por un proceso semejante. No resulta plausible que dispongamos de ejemplares, ni prototipos, ni tampoco de conocimiento teórico de la categoría de las abuelas la mayoría de cuyos nietos son dentistas. En este sentido, este

concepto no puede explicarse por la combinación de los rasgos de los prototipos de ABUELA, NIETO y DENTISTA. Podría haber raras excepciones de individuos que tengan de hecho uno o varios ejemplares de la categoría y puedan abstraer un prototipo a partir de ellos. Pero, en general, en los casos como este, se puede recurrir a la composicionalidad de etiquetas simbólicas.

Machery (2009) ofrece un modelo similar al de Prinz para la combinación de conceptos, en el que los distintos tipos de estructuras conceptuales interactúan para formar un concepto compuesto. Machery explica la combinación de conceptos como la combinación de rasgos provenientes de distintos tipos de estructuras semánticas. Así, combinar conceptos sería un proceso creativo que toma algunos de los rasgos de cada concepto interviniente. Se pueden tomar rasgos de prototipos y de ejemplares, y “editar” el conjunto de rasgos final en función del conocimiento teórico acerca de las categorías representadas por los conceptos que se quiere combinar. De este modo, a diferencia de la categorización, en la que se pueden identificar distintos tipos de procesos, la combinación se llevaría a cabo por un único proceso que toma distintos tipos de *input*.

Así, hay diversos modelos disponibles que resultan compatibles con las tesis del pluralismo amplio, al menos en lo relativo a los distintos tipos de estructuras semánticas. Me ocuparé de la interacción entre los distintos tipos de vehículos en §4-§6.

Adquisición

A diferencia de otras propuestas pluralistas (Weiskopf 2009a), considero plausible que cada tipo de estructura requiera un proceso distinto de adquisición. Por ejemplo, un encuentro con una instancia de GATO puede resultar suficiente para adquirir un ejemplar, pero seguramente no resultará suficiente para formar un prototipo, para lo cual se requieren varias representaciones a partir de las cuales se puedan abstraer las características generales o tendencias centrales de la categoría, así como mecanismos inductivos de generalización. Podría ser que los prototipos se adquieran por abstracción o generalización a partir de conjuntos de ejemplares almacenados. En este sentido, un tipo de concepto puede intervenir en la adquisición de otros tipos de conceptos. Esto podría contribuir a resolver el problema de Fodor respecto de la relación entre prototipos y etiquetas, dado que en lugar de tratarse de un pasaje de representaciones no-conceptuales a representaciones conceptuales, sería un pasaje entre tipos de conceptos.

En este sentido, el hecho de postular representaciones perceptivas conceptuales acarrea una ventaja para explicar la adquisición de conceptos, puesto que ofrece un conjunto

de representaciones que son el *output* de sistemas de *input* dedicados innatos y pueden contribuir en la adquisición de otros conceptos. Carey (2009) ofrece buenos ejemplos de cómo conceptos vehiculados por representaciones perceptivas pueden contribuir en la adquisición de otros conceptos. Uno de estos ejemplos es el de las representaciones de números, que veremos en §5. Me ocuparé del vínculo entre los distintos tipos de vehículos las próximas secciones, al sugerir qué tipo de arquitectura resulta compatible con el pluralismo amplio.

Por último, no es necesario comprometerse, como hace Fodor, con que la adquisición de los prototipos sea una fase inicial en la adquisición de conceptos, en tanto, por un lado, adquirir un prototipo sería ya adquirir un concepto y, por otro lado, no veo la necesidad de adquirir un prototipo antes que un símbolo amodal. El mecanismo de acceso semántico al mundo puede estar dado por un conjunto de representaciones perceptivas que no resulten suficientes para abstraer un prototipo, pero sí para abrir un archivo relacionado con una categoría y etiquetarlo con un símbolo. La adopción de un enfoque pluralista permite sostener que se ha adquirido un concepto cuando se abre un archivo para almacenar las representaciones de una categoría en particular y las primeras representaciones que se almacenan pueden ser conceptos de cualquiera de las estructuras que el pluralismo incorpora, siempre y cuando resulten suficientes para realizar procesos cognitivos superiores. Así, cualquier tipo de concepto sirve para abrir un archivo, en el sentido en que sirve como primera representación de una propiedad, aunque considero que la apertura de un archivo debe ir acompañada siempre de la asignación de una etiqueta al mismo, que es la que individualiza el archivo y fija su referencia.

Alcance

Este requisito no debería resultar problemático en absoluto para el pluralismo. El enfoque que propongo no exige que para cada categoría dispongamos de una definición, un prototipo, un conjunto de ejemplares y una teoría, aunque sí una etiqueta (una representación simbólica atómica) que marca la apertura de cada archivo y fija su referencia. De este modo, los conceptos abstractos y teóricos de los cuales no se puede dar cuenta apelando a ejemplares ni a representaciones perceptivas, no acarrear dificultades, puesto que pueden ser representados por medio de símbolos atómicos amodales.

La inclusión de definiciones, a su vez, permite extender el alcance a conceptos lógicos y matemáticos. Teniendo en cuenta las consideraciones de Fodor (1999), parecería que la mejor manera de dar cuenta de los conceptos de CONJUNCIÓN o RECTA es por medio de

definiciones. Como vimos, otros autores también afirman que las definiciones resultan adecuadas para otros conceptos como los de parentesco. Si bien el alcance de las definiciones resulta limitado, parecen adecuadas para algunos conceptos en particular. No solo esto, sino que parecen más adecuadas que el resto de las estructuras semánticas para dar cuenta de la representación de estos tipos particulares de conceptos.

De este modo, el pluralismo amplio cuenta con los recursos representacionales para dar cuenta del amplio espectro de conceptos con los que nuestra mente puede operar, desde los más concretos hasta los más abstractos, teniendo en cuenta tanto aquellos a cuyas instancias tenemos acceso perceptivo como aquellos para los que éste no es el caso.

Publicidad

El pluralismo amplio puede suscitar objeciones respecto de las posibilidades de satisfacer el requisito de publicidad. Consideraré dos objeciones posibles. Primero, el hecho de incorporar distintos tipos de estructura semántica interna, introduce probablemente diferencias entre los individuos. En particular para el caso de los ejemplares, como sostuve en el capítulo 2. Dado que son conjuntos de representaciones de instancias particulares de una categoría con las que estuvimos en contacto, no es probable que distintos individuos compartan tales conjuntos. Así, los ejemplares no satisfacen el requisito de publicidad. En consecuencia, el pluralismo amplio postula por lo menos un tipo de estructura que no es compartida. Sin embargo, postula también representaciones atómicas, que serían las mismas en tipo para distintos individuos. Si el modelo de articulación del contenido que propuse es adecuado, entonces los distintos tipos de estructuras semánticas estarían asociados a un mismo archivo y a una representación atómica compartida. En este sentido, si bien es posible que no se compartan todas las estructuras semánticas que representan una categoría, el símbolo que la individúa sí es compartido.

Recordemos las razones para postular el requisito de publicidad. Éste se postula, primero, para que sean posibles las generalizaciones psicológicas para explicar la conducta. Éstas serían generalizaciones sobre el contenido de las creencias y los deseos, de modo que, para que sean posibles, tales contenidos tienen que ser compartidos. Segundo, la publicidad permite dar cuenta de la comunicación lingüística, bajo el supuesto de que los conceptos son los significados de las palabras. Si el significado de lo que expresamos lingüísticamente está estructurado en conceptos y no compartimos conceptos, entonces la comunicación lingüística se torna un misterio. Pues bien, la postulación de un conjunto de representaciones conceptuales que pueden no ser compartidas, torna la medida en la que compartimos

conceptos en una cuestión de grado. Las diferencias estarían dadas por el contenido cognitivo por medio del cual distintos individuos representan una categoría.

Imaginemos que dos personas están discutiendo acerca de la posibilidad de que se extraiga ADN de dinosaurio de una roca de ámbar y mediante ciertos procedimientos en un laboratorio se obtenga un dinosaurio vivo. Es posible que el contenido cognitivo que asocian a ADN, DINOSAURIO, ÁMBAR, etc. sea diferente. El hecho de compartir representaciones atómicas para cada una de esas categorías, que están causalmente conectadas con las propiedades que representan, asegura que están hablando de lo mismo y que, en cierta medida, se comprenden mutuamente. Pero si una de ellas tiene una teoría científica del ADN, con detalles respecto de su composición y sus propiedades, y la otra una teoría ingenua, no identificarán esta categoría por medio de las mismas representaciones. Esto podría dar lugar a ciertas limitaciones en la comunicación. Ahora bien, esto no necesariamente es una desventaja. La comunicación lingüística es un fenómeno estable que requiere que compartamos significados, pero también se dan, ocasionalmente, casos de incomunicación. No siempre nos entendemos. Y esto es algo que se puede explicar por las diferencias en el contenido cognitivo por medio del cual representamos cada categoría.

Una segunda objeción posible se relaciona con la individuación de archivos. Al desarrollar la teoría semántica que adopta el pluralismo amplio sostuve que adhería a una semántica de doble factor. Los conceptos están vinculados con propiedades por medio de relaciones nomológicas que les confieren contenido y, además, tienen una estructura interna que constituye su contenido cognitivo y que refleja el modo en que concebimos las categorías. Con respecto a la objeción previa sostuve que compartimos etiquetas atómicas, dado que éstas no tienen contenido cognitivo, sino únicamente el contenido dado por relaciones nomológicas con propiedades. Sin embargo, también sugerí que en la individuación de archivos intervienen factores epistémicos, resultando de especial relevancia el hecho de que creamos que distintas representaciones son acerca de la misma categoría. Así, el modo en que se agrupan los distintos tipos de estructuras semánticas en cada archivo es una cuestión epistémica. Dado que, en principio, cualquier creencia podría intervenir en la organización conceptual y que es poco probable que distintos individuos compartan sistemas completos de creencias, esta propuesta parece conducir al holismo.

Si bien este riesgo está presente, hay ciertas estrategias disponibles para adoptar. Mencionaré aquí una respuesta posible, relacionada con la mención de “en principio” para señalar que cualquier creencia podría intervenir en la individuación de conceptos y, en

consecuencia, ésta depende del sistema completo. Keil y Wilson (2000) ofrecen una analogía con el sistema solar para ilustrar por qué éste no es el caso:

Todas las lunas de nuestro sistema solar están influenciadas en sus órbitas por todo el resto de las masas del sistema solar; pero cada planeta y sus lunas forma un sistema coherente distinto de cualquier otro, constituyendo un sistema que puede ser prácticamente comprendido por completo en ese nivel de análisis. (p. 314)

Adoptando la misma estrategia que se sugiere en la analogía, las creencias que permitirían determinar qué representaciones se asocian a cada archivo, i.e. qué estructuras semánticas se consideran correferenciales, podrían estar fuertemente interconectadas y, en cierta medida, poder aislarse del resto del sistema de creencias. De este modo, la posibilidad de compartir conceptos resulta, por lo menos, plausible. Habrá ciertas diferencias en las estructuras semánticas particulares que distintos individuos asocian a un archivo u otro, pero, por lo menos es probable que la organización en archivos en general sea compartida.⁵⁶

De este modo, la evaluación de cada uno de los requisitos a partir de la tesis que postula una pluralidad de estructuras semánticas sugiere que este tipo de propuesta presenta ventajas explicativas en relación a los distintos requisitos fundamentales. A continuación abordaré la tesis relativa a la pluralidad de vehículos conceptuales e intentaré mostrar que ésta permite dar cuenta de la evidencia disponible, así como complementar la explicación de algunos de los requisitos ofrecida hasta aquí.

3. Tipos de vehículos conceptuales y Tipos de procesamiento

Desde el surgimiento de la ciencia cognitiva, ha estado entre sus objetivos la caracterización de la arquitectura cognitiva, i.e. la descripción general de los recursos y la organización estructural subyacentes a la mente (Newell 1990). La descripción de una arquitectura cognitiva pretende dar cuenta de las características generales, y la organización de los recursos mentales (tales como la memoria), y los procesos mentales. Así, pretende especificar, por ejemplo, si los procesos son seriales o pueden operar en paralelo, si hay distintos vehículos representacionales o un único tipo de vehículo, y si hay procesos modulares o no. Como defendí en los capítulos 4 y 5, hay evidencia y argumentos que

⁵⁶ Por supuesto que habrá casos en los que esto no sea así. Pero el requisito es modal, no requiere que todos los conceptos sean compartidos, sino que sean compatibles, en el sentido de no ser privados.

sugieren que no hay un único tipo de vehículo conceptual, sino que los conceptos pueden estar vehiculados tanto por representaciones de tipo simbólico como de tipo perceptivo. Esto presupone ciertas características arquitecturales, puesto que los distintos tipos de vehículos implican distintos tipos de procesos y operaciones en los que cada uno interviene.

Como vimos en el capítulo 4, sobre las representaciones amodales pueden operar transiciones causales guiadas por la forma, con el caso paradigmático de las inferencias lógicas. Los procesos de adquisición de representaciones amodales involucran operaciones de transducción desde los sistemas perceptivos o de *input* y cada uno de sus formatos específicos, a una suerte de código común de la mente en el que se expresan los conceptos, el LDP. Según los defensores de estas representaciones, los procesos de pensamiento conceptual consisten en transiciones causales entre símbolos, sensibles a la forma de las representaciones y que preservan el contenido (Fodor 1975, 1983, 2008). Los procesos de adquisición de representaciones modales, en cambio, involucran el registro multimodal de patrones de activación que pasan a estar bajo control del organismo para su posterior reactivación, como vimos en el capítulo 5. Con respecto a los procesos de pensamiento, se trata de simulaciones o reactivaciones de estados perceptivos, en tanto consisten en la reactivación de los registros multimodales producidos a partir de interacciones con el entorno (Barsalou 1999, Prinz 2002).

Dada la evidencia y los argumentos presentados a favor de estos dos tipos de vehículos, sugerí que el pensamiento conceptual puede estar vehiculado en dos formatos representacionales diferentes y que, en consecuencia, hay (por lo menos) dos tipos de procesamiento conceptual distintos: el pensamiento simbólico y la simulación. Por otra parte, también habría distintos tipos de procesos en función de cada tipo de estructura semántica. Por ejemplo, la categorización por medio de prototipos o ejemplares consistiría en una función que compara similitudes entre rasgos y determina la pertenencia a una categoría si se supera cierto umbral (Hampton 1995, Smith & Osherson 1989, Tversky 1977). La categorización por medio de teorías, en cambio, consistiría en una inferencia a la mejor explicación que asigna una instancia a la categoría cuyos principios causales explican mejor los rasgos de la instancia en cuestión (Rips 1989). Esta multiplicidad de procesos sugiere que la mente cuenta con distintos sistemas, cada uno identificado por un tipo de representaciones y procesos característicos. En este capítulo propongo que a nuestras capacidades conceptuales subyacen dos tipos principales de sistemas, uno modal y otro amodal. En este sentido, propongo adoptar (y adaptar) una versión de arquitectura dual.

Las arquitecturas duales han ganado creciente atención en los últimos años, en distintos ámbitos de la ciencia cognitiva, por sus ventajas explicativas respecto de un amplio espectro de fenómenos. Estas propuestas se desarrollaron para explicar capacidades diversas como el razonamiento (Evans 2003, Sloman 1996, Stanovich 2004, Tversky & Kahneman 1983), los juicios morales (Cushman & Greene 2012, Cushman, Young & Hauser 2006, Greene 2009, Greene, Nystrom, Engell, Darley & Cohen 2004), el aprendizaje (Dienes & Perner 1999, Reber 1993, Sun, Slusarz & Terry 2005), la cognición numérica (Carey 2009, Feigenson, Dehaene & Spelke 2004), la categorización (Sloman 1996, 2002), la cognición social (Lieberman 2007) y la capacidad de *mindreading* (Apperly & Butterfill 2009). Las distintas propuestas coinciden en postular que a cada una de esas capacidades subyacen dos tipos de sistemas o procesos, cada uno con características distintivas, y suelen coincidir a grandes rasgos en las propiedades atribuidas a cada uno de los sistemas. Sin embargo, no hay una caracterización única de las propiedades de cada uno de ellos.

En general, las distintas propuestas se han desarrollado para explicar fenómenos que parecen revelar una competencia o conflicto entre dos sistemas relativamente independientes, que permiten llevar a cabo una determinada capacidad cognitiva de modos distintos. Un ejemplo clásico es el de la falacia de conjunción que se observa en los casos en los que individuos con entrenamiento en estadística atribuyen mayor probabilidad a una conjunción que a una de sus partes. En un estudio de Tversky & Kahneman (1983) se presenta a Linda, con las siguientes características:

Linda tiene 31 años, es soltera, franca y muy inteligente. Se especializó en filosofía. Como estudiante, estaba profundamente preocupada por los problemas de discriminación y justicia social, participando también en manifestaciones anti-nucleares. (p. 297)

En el estudio los participantes debían leer la presentación de Linda y luego evaluar la probabilidad de una serie de afirmaciones. Los resultados revelan que los individuos asignan una probabilidad mayor a que Linda sea cajera en un banco y activista en movimientos feministas, que la que asignan a que Linda sea cajera en un banco. Así, violan una regla o principio de los razonamientos probabilísticos por la cual la probabilidad de una conjunción no puede exceder la probabilidad de sus constituyentes.

Cada ámbito de investigación para el que se ha propuesto una arquitectura dual presenta fenómenos específicos que requieren explicación. Sin embargo, en general todos

ellos revelan la aplicación de conjuntos de principios distintos, que conducen a resultados inconsistentes, como en el ejemplo de Tversky y Kahneman (1983). Para explicar este tipo de fenómenos, las teorías de arquitecturas duales sugieren que disponemos de dos sistemas relativamente independientes, que subyacen a una determinada capacidad cognitiva y pueden conducir a patrones de desempeño distintivos. Estos sistemas serían sólo relativamente independientes dado que, como veremos más adelante, puede haber interacción e interferencia entre ellos. Algunos autores se refieren a estos sistemas como “dos mentes”, por las diferencias entre ellos (e.g. Sloman, 2002). Los dos tipos de procesos postulados se han denominado de distinto modo dependiendo de la propuesta y, a la vez, se han sugerido distintos conjuntos de rasgos asociados a cada tipo de proceso.

Las caracterizaciones principales propuestas en la literatura distinguen entre: (i) procesos implícitos que son automáticos, rápidos y operan en paralelo, y procesos explícitos que son controlados, secuenciales y lentos (Reber, 1993; Sun, 2002); (ii) procesos modulares, de dominio específico, obligatorios, informacionalmente encapsulados, rápidos, que arrojan resultados superficiales y exhiben patrones de disfunción y de desarrollo ontogenético específicos, y procesos centrales, de dominio general, que no están informacionalmente encapsulados (Fodor, 1983); (iii) procesos asociativos, que reflejan similitud y estructura temporal y codifican regularidades estadísticas, y procesos basados en reglas, que operan sobre representaciones simbólicas, son secuenciales y satisfacen los principios computacionales de productividad y sistematicidad (Sloman 1996); (iv) procesos automáticos y computacionalmente poco demandantes que consumen pocos recursos tales como la memoria de trabajo, y procesos que están bajo control interno del organismo y son altamente demandantes de recursos computacionales (Evans & Stanovich 2013); y (v) procesos intuitivos que son procesos de reconocimiento de patrones en “un medio representacional que integra estructuras de conocimiento con procesos físicos y produce tanto *outputs* simbólicos como estados afectivos” (Sloman 2014, p. 75), y procesos deliberativos que son procesos de pensamiento reflexivo que operan sobre un medio simbólico (Sloman 2014). Algunos autores identifican también los procesos del primer tipo (ya sea que se caractericen como implícitos, modulares, asociativos o intuitivos) como propios de sistemas evolutivamente antiguos, y explican así el hecho de que compartimos ciertas capacidades con otras especies, mientras que los procesos del segundo tipo corresponderían a sistemas evolutivamente recientes, vinculados al lenguaje y distintivamente humanos (Carey 2009, Evans 2003).

Esta breve mención de las principales características atribuidas a los dos tipos de procesos muestra algunas de las diferencias entre las distintas posturas. La proliferación de

distinciones que se superponen sólo parcialmente ha suscitado críticas que señalan la imposibilidad de encontrar una caracterización homogénea de los dos tipos de procesos que sea común a todas las teorías de procesos duales (Keren & Schul 2009, Kruglanski & Gigerenzer 2011). Sin embargo, como señalan Evans y Stanovich (2013), esta crítica sólo tendría sentido si las características de los sistemas se consideraran definitorias, i.e. necesarias y suficientes para constituir alguno de los dos sistemas. Generalmente, los autores caracterizan uno y otro tipo de procesos por una serie de rasgos típicamente correlacionados, y no por medio de rasgos definicionales. Sin embargo, a fin de establecer las tesis básicas de la arquitectura cognitiva, compartidas por las distintas teorías, es preciso identificar por lo menos una característica que permita distinguir los dos tipos de procesos. De otro modo, no tendría sentido atribuir alguna tesis común a las distintas propuestas de arquitectura dual. Stanovich y Toplak (2012) señalan que la identificación de una única característica necesaria y suficiente alcanza para distinguir los dos sistemas o tipos de procesos, mientras que el resto de las características pueden considerarse como típicamente correlacionadas con cada sistema, pero no definitorias.

Otro tipo de objeción a este tipo de teorías señala que a las capacidades cognitivas subyace un continuo de procesos distintos, en el cual no es posible establecer una distinción discreta entre dos tipos principales de procesos, tal como la pretendida por las arquitecturas duales (Osman 2004). De este modo, si bien podrían diferenciarse distintos procesos subyacentes a una misma capacidad, adoptando así teorías de procesos múltiples, estos no corresponderían a dos sistemas distintos, sino a múltiples sistemas (Machery 2009). Esta crítica está íntimamente relacionada con la anterior, puesto que señala la imposibilidad de identificar, en un continuo de procesos que comparten en mayor o menor medida una serie de características, alguna nota central que permita individuar dos tipos de procesos correspondientes a dos sistemas cualitativamente distintos.

Considero, siguiendo a Sloman (1996, 2002, 2014), que los dos sistemas se pueden distinguir en función del tipo de vehículo sobre el que opera cada uno, así como el tipo de procesos que caracterizan a uno y otro. Sloman propone distinguir entre procesos intuitivos y procesos deliberativos. Los primeros son de tipo asociativo, registran regularidades estadísticas y establecen generalizaciones a partir de ellas, así como comparaciones de similitud entre situaciones o instancias nuevas y experiencias pasadas. Los segundos, en cambio, consisten en procesos de transiciones formales o aplicación de reglas formales, que se basan en principios lógicos y relaciones causales. Sloman postula que ambos tipos de procesos involucran medios representacionales distintos: los primeros operarían sobre redes

de representaciones asociadas (patrones de rasgos asociados), mientras que los segundos operarían sobre representaciones simbólicas. Si bien no establece una distinción explícita entre tipos de vehículos, Sloman (1996) sostiene que el sistema deliberativo opera sobre representaciones de tipo simbólico como el LDP (Fodor 1975), que vimos en el capítulo 4.

Sloman menciona los argumentos de Fodor y Pylyshyn (1995), reconstruidos en el capítulo 1, que sostienen que la composicionalidad es la mejor explicación para la sistematicidad y productividad del pensamiento, y por ende que el pensamiento tiene que estar vehiculado en un medio representacional simbólico de modo de satisfacer el requisito de composicionalidad. Sloman adhiere a tales conclusiones, pero señala que su alcance debe restringirse a un tipo de razonamiento y, de modo más general, de pensamiento: el deliberativo. Sin embargo, señala que ésta no es la única forma de razonamiento, sino que disponemos también de razonamiento intuitivo. En términos de Sloman, el medio representacional del sistema intuitivo difiere del simbólico puesto que opera sobre procesos físicos (Sloman 2014). Dadas estas características, sugiero interpretar el sistema intuitivo como un sistema modal, entendiendo los procesos físicos como estados perceptivos. Recordemos que tal como postulan Barsalou (1999) y Prinz (2002), los sistemas conceptuales de representaciones modales se basan en registros de la activación simultánea de representaciones en distintas modalidades y en la activación posterior de tales registros de patrones estadísticos. Si bien involucran mecanismos atencionales que pueden influir en los registros, se trata, fundamentalmente, de mecanismos de detección de correlaciones estadísticas, del mismo tipo que los propuestos por Sloman para el sistema intuitivo.⁵⁷

Sloman (1996) toma de la literatura de conceptos un ejemplo que propone explicar a partir de la distinción arquitectural entre el sistema intuitivo y el deliberativo. Sloman apela a parte de la evidencia presentada en los capítulos 2 y 3, relativa a los distintos tipos de efectos asociados a tareas de categorización: los efectos de tipicidad y los efectos causales. Hay

⁵⁷ Sloman (1996) sugiere que la diferencia entre los dos tipos de vehículos (o medios representacionales, como él los denomina) puede modelarse como la diferencia entre una red conexionista de representaciones subsimbólicas, y un sistema de símbolos clásicos. Una red conexionista es un modelo artificial que consiste en una red de unidades interconectadas. La fuerza de las asociaciones entre las unidades se modela asignándoles un peso, que puede aumentar o disminuir en función de la frecuencia de activación conjunta de las distintas unidades. Un sistema clásico, en cambio, es un sistema como el LDP con un conjunto de símbolos que poseen una sintaxis y una semántica combinatoria y sobre los que pueden operar transiciones causales dependientes de la forma y que preservan el contenido (Fodor 1975, 1987). No es mi intención asociar los sistemas modales de representaciones con las redes conexionistas, sino simplemente señalar que hay una superposición significativa entre las características atribuidas al sistema intuitivo y las atribuidas a los sistemas de representaciones modales y que, en este sentido, se puede interpretar la diferencia de medio simbólico como la diferencia entre representaciones modales y amodales. Los defensores de los vehículos conceptuales modales suelen aludir a modelos de redes conexionistas como posibles implementaciones de las tesis que defienden (Barsalou 1999, Pecher 2013), por lo que proponer que el sistema intuitivo opera sobre representaciones modales parece adecuado.

evidencia del rol de la similitud en la categorización, por la ubicuidad de los efectos de tipicidad, tales como la velocidad con la que categorizamos instancias típicas de una categoría. Pero, a la vez, en algunos casos se categoriza una instancia como perteneciente a una categoría con la que guarda menos similitud que con otra, como en el caso del ejemplo de Rips (1989), que vimos en el capítulo 3, en el que se debía determinar si algo circular de un determinado diámetro era una pizza o una moneda.

Sloman presenta esta evidencia para afirmar que en los dos efectos se llevan a cabo procesos de categorización de distinto tipo. En un caso categorizamos por medio de procesos de comparación de conjuntos de rasgos, en los que se computa la fuerza de asociación entre los rasgos en función de correlaciones estadísticas. Este tipo de procesos corresponde al sistema intuitivo. En el otro caso, en cambio, construimos una explicación, por medio de un proceso deliberativo. El objetivo de tal explicación es que la atribución de una instancia a una categoría nos permita explicar sus rasgos.⁵⁸ En esta caracterización se refleja otra diferencia entre los dos tipos de procesos. Mientras que en el caso de los procesos intuitivos, tenemos acceso sólo al resultado, los procesos deliberativos, en cambio, son accesibles por el sujeto (Sloman 2014).

La distinción entre sistemas en términos del acceso a los procesos parece aplicarse a la categorización. Cuando categorizamos por medio de prototipos y ejemplares no tenemos acceso a los procedimientos de comparación de rasgos, sino sólo a sus resultados, mientras que la categorización por medio de teorías es un proceso dirigido intencionalmente por el sujeto, al cual tiene acceso en sus distintos pasos. Sin embargo, esto no significa que resulte adecuada para todos los procesos conceptuales. Esta distinción no tiene sentido para, por ejemplo, los procesos de adquisición de conceptos. No resulta plausible que tengamos acceso a los procesos por los que las etiquetas atómicas adquieren contenido, del mismo modo que no necesariamente formulamos conscientemente hipótesis acerca de las relaciones causales o los principios subyacentes a las categorías biológicas (probablemente esto dependa de la categoría en cuestión). Dado que pretendo extender la aplicación de los dos sistemas propuestos por Sloman a los conceptos en general, y dadas las características en común entre el sistema intuitivo y los sistemas modales, por un lado, y entre el sistema deliberativo y los sistemas amodales, por otro, me referiré directamente al primero como sistema intuitivo o modal y al segundo como sistema deliberativo o amodal. Así, estas denominaciones recogen más claramente una de las distinciones principales del pluralismo amplio.

⁵⁸ En ocasiones Sloman se refiere al segundo tipo de proceso como “gobernado por reglas”. Esto no significa que los procesos del primer tipo no sigan reglas, sino que no son reglas que dependen de la forma.

De este modo, si bien las características de estos dos sistemas se propusieron específicamente para dar cuenta del razonamiento, podrían dar cuenta también de la organización de nuestro sistema conceptual. Sloman tiene en cuenta fenómenos comúnmente abordados en los debates en torno a los conceptos de los que me he ocupado a lo largo de la tesis (como la atribución de propiedades basada en la correlación estadística entre rasgos, o basada en un razonamiento que apela a relaciones causales), lo cual sugiere un vínculo entre su propuesta y las teorías de conceptos. Como evidencia la reconstrucción presentada, la propuesta de Sloman es consistente con el pluralismo amplio, dado que postula dos sistemas que operan sobre representaciones con distintos tipos de vehículos. Ahora bien, teniendo en cuenta la caracterización ofrecida de definiciones, prototipos, ejemplares, teorías y etiquetas atómicas, así como de los vehículos modales y amodales, ¿cómo se distribuirían estas representaciones entre los dos sistemas? En la próxima sección ofreceré una respuesta y delinearé la versión de organización de los procesos conceptuales que emerge al asociar las propuestas de la arquitectura dual y el pluralismo amplio.

4. El pluralismo amplio y una arquitectura dual

Hasta ahora no me comprometí con ninguna compatibilidad específica entre los tipos de estructuras semánticas y los tipos de vehículos. Primero, es importante aclarar que esta será, en última instancia, una cuestión empírica, que debe ser zanjada mediante estudios psicológicos o neuropsicológicos. Sin embargo, se pueden hacer ciertas sugerencias especulativas respecto de qué vehículo resulta más compatible con cada estructura semántica a partir de los modelos de procesamiento que proponen los principales defensores de cada tipo de estructura.

Primero, las etiquetas o símbolos atómicos son amodales. De modo que estas representaciones corresponderían al sistema deliberativo. Dadas las consideraciones de Sloman relativas a la sistematicidad y productividad del sistema deliberativo, resulta natural sostener que este sistema opera sobre representaciones atómicas amodales. Las teorías presumiblemente también están vehiculadas por símbolos amodales. Algunas de las razones para afirmar esto están dadas por la discusión con la teoría de proxitipos, presentada en el capítulo 5. Prinz afirma que las representaciones perceptivas pueden portar información de distinto tipo, incluida la información de tipo teórico. Sin embargo, como argumenté en relación a las dificultades de los proxitipos para dar cuenta del alcance, no parece posible representar relaciones causales (que son centrales en las estructuras de las teorías) por medio de representaciones perceptivas, sin introducir representaciones de otro tipo. Probablemente

Prinz deba recurrir para este tipo de representaciones a palabras del lenguaje natural y, como sostuve, estaría introduciendo así un sistema de representaciones amodales.

Además, se puede señalar que los procesos que postulan los defensores del enfoque de teoría-teoría no son de tipo asociativo, sino que son, como señala Sloman, característicos del sistema deliberativo. Se trata, como vimos en el capítulo 3, de inferencias a la mejor explicación que atribuyen a una instancia la categoría que mejor explica sus atributos. Estos procesos no dependen de la similitud ni de la asociación, sino de relaciones causales y explicativas entre los rasgos. Por último, las definiciones también corresponderían al sistema deliberativo, dado que los procesos que operan sobre ellas consisten en la aplicación de reglas y no dependen de asociaciones estadísticas.

Los prototipos, a su vez, estarían en el dominio del sistema modal, por el hecho de que los procesos que operan sobre ellos son de tipo asociativo y de comparación de patrones o de rasgos superficiales. Lo mismo valdría para los conjuntos de ejemplares. Si se prioriza el criterio del tipo de procesamiento, los prototipos y los conjuntos de ejemplares corresponderían al sistema modal. Esto puede resultar extraño, puesto que los modelos ofrecidos por los defensores de estas teorías no parecen atribuirles las características de las representaciones perceptivas. Sin embargo, la identificación no parece descabellada a la luz de una serie de consideraciones que presento a continuación.

Primero, algunos de los principales modelos de procesamiento de ejemplares y prototipos recurren a redes conexionistas (Nosofsky 1992). Esto coincide con el modo en que Sloman caracteriza los procesos del sistema intuitivo. Segundo, Rosch y sus colegas (1976) se refieren a las categorías de nivel básico como “el nivel taxonómico más inclusivo respecto del cual es posible tener una representación que sea isomorfa con la apariencia física de los objetos de la clase” (p. 406). Si bien no se refieren a los prototipos como realizados en representaciones isomorfas con aquello que representan, destacan que las categorías de nivel básico tienen este tipo de representaciones asociadas y, como vimos en el capítulo 5, las representaciones modales son en cierta medida isomorfas con aquello que representan.⁵⁹ Por último, para resolver tareas de verificación de propiedades, tales como las que se presentan como evidencia del uso de sistemas perceptivos, es plausible que los tipos de representaciones que se activan sean prototipos *por default*, del tipo que postula Prinz (2002), porque en estas tareas no se ofrece ningún contexto específico.

⁵⁹ Las categorías de orden inferior al nivel básico pueden tener también tales representaciones asociadas. Las de orden superior, en cambio, no, dado que las instancias a las que refieren son tan diversas que “no puede haber una representación isomorfa con los rasgos físicos de la categoría como un todo” (Rosch *et al.* 1976).

En la próxima sección presentaré un caso de aplicación de este tipo de arquitectura, el caso de los números naturales, que me servirá para ilustrar el tipo de articulación entre los dos tipos de sistemas que considero adecuada.

5. El caso de los conceptos de números

Nuestras habilidades numéricas intervienen en gran parte de nuestra vida mental. No sólo se ponen en práctica al contar, obtener la cardinalidad de un conjunto, comparar valores de una determinada magnitud o llevar a cabo operaciones aritméticas, sino que también están involucradas en el aspecto cuantificacional del lenguaje. El estudio de estas habilidades, su base representacional y neuronal, así como su desarrollo, constituye un campo de investigación en sí mismo. Mi intención aquí es presentar brevemente una de las teorías del desarrollo de los sistemas conceptuales numéricos que apela a una arquitectura dual, elabora una articulación entre un conjunto de representaciones iniciales que son producidas por sistemas innatos específicos de dominio y sugiere que mecanismos dan lugar a representaciones más complejas que las iniciales. Esta propuesta resulta interesante en relación al pluralismo amplio, dado que adopta una versión de la adquisición de conceptos similar en algunos aspectos a la propuesta por los defensores de los vehículos modales, pero incorpora mecanismos que permiten adquirir representaciones en un vehículo distinto al perceptivo y elucida el vínculo entre conceptos con distintos vehículos. Por esta razón, sugeriré que el pluralismo amplio es compatible con una propuesta de este tipo, que ha resultado útil en otros dominios.

Esta propuesta adhiere a la tesis de que hay discontinuidades conceptuales entre los niños y los adultos en algunos dominios, pero no postula una discontinuidad radical, sino que sostiene que también tienen recursos conceptuales en común (Carey 2004, 2009, Mandler 2004). Como sostiene Carey, para el caso de los números, en particular, los niños preverbales exhiben una serie de habilidades numéricas que se pueden explicar postulando dos sistemas de representación: un sistema que representa magnitudes analógicas y un modelo de individuación de objetos. El primero es el que permite explicar la discriminación entre valores, e.g. entre conjuntos que tienen más o menos elementos. Este sistema está sujeto a una limitación que se conoce como la *ley de Weber*, según la cual la discriminabilidad entre dos valores es una función de su razón. Cuanto menor sea la razón entre los valores, más difícil resulta discriminarlos (Dehaene 1997).

Se ha observado esta capacidad en niños de a partir de 6 meses, utilizando paradigmas de habituación, en los que se presenta un estímulo a niños pequeños hasta que se habitúan a él

y luego presentan otro estímulo. Lo que se mide en estos estudios es si recuperan la atención ante el nuevo estímulo o no. Frente a conjuntos de puntos, se observó que, habituados a 8 puntos, los niños recuperan la atención ante 16, pero no ante 12, lo cual sugiere que discriminan entre 8 y 16, pero no entre 8 y 12. De la misma manera, discriminan entre 4 y 8, y entre 16 y 32, pero no entre 16 y 24 (Xu & Spelke 2000). Estudios similares se llevaron a cabo, por ejemplo, con tonos (Carey 2009).

El segundo sistema es el que permite individuar conjuntos pequeños de objetos (hasta tres) y rastrearlos en el tiempo (aunque sin por esto individuarlos en tanto instancias de tal o cual categoría). Este sistema asigna un símbolo a cada individuo del conjunto y exhibe un límite superior relativo al tamaño del conjunto que puede individuar, dado que permite individuar hasta tres elementos. Por ejemplo, se han llevado a cabo estudios en los que niños de 12 a 14 meses observan cómo alguien pone objetos en una caja (uno a la vez) y después pueden buscar en la caja un objeto a la vez, pero no pueden ver el interior. Cuando observan que se ponen en la caja uno, dos o tres objetos, los niños buscan exactamente la cantidad observada. En cambio, cuando se supera el límite de tres objetos, por ejemplo si se ponen cuatro objetos en la caja, los niños se contentan con sacar un objeto de la caja, o en algunos casos dos, como si hubieran perdido la capacidad de rastrear los objetos (Feigenson & Carey 2003).

Si bien estos sistemas representan magnitudes numéricas y permiten, en consecuencia, ejecutar una serie de habilidades numéricas, se diferencian del sistema de representación de números naturales. El sistema de representación de magnitudes analógicas, primero, no permite representar cantidades exactas, sino sólo de modo aproximado, segundo, está sujeto a la ley de Weber y, tercero, no representa la función sucesor (i.e. la función por la cual dos números sucesivos se diferencian en una unidad, o el siguiente de un número es ese número más uno). El sistema de individuación de objetos, por otra parte, no incluye representaciones explícitas de números, sino que individúa de un modo implícito la cardinalidad de un conjunto, y lo hace hasta un límite superior de tres elementos (Carey 2004, 2009).

El hecho de que estas habilidades se manifiesten en niños muy pequeños (desde aproximadamente los 6 meses), ha llevado a postular que éstas se explican por un conjunto de representaciones conceptuales innatas, que se denomina usualmente “cognición nuclear” (Carey & Spelke 1994, Carey 2009). Se han propuesto este tipo de representaciones para los dominios de los objetos, agentes y números. La cognición nuclear incluiría, además de un conjunto de representaciones, principios universales sobre la estructura y entidades de un dominio. Como vimos en el capítulo 3, estos ayudan a resolver el problema de la postulación

de hipótesis relativas al contenido de los conceptos, así como a dirigir la atención selectivamente hacia determinadas propiedades o aspectos de las categorías.

Se puede interpretar que las representaciones de la cognición nuclear son modales en un sentido similar al que proponen Barsalou (1999) y Prinz (2002): son *outputs* de sistemas perceptivos innatos analizadores de *input* (Carey 2009). Más aun, Carey se refiere al vehículo de estas representaciones como icónico. Al igual que Fodor (2008) al diferenciar representaciones icónicas y simbólicas, Carey sostiene que las partes de las representaciones icónicas representan partes de la entidad representada. Una parte de una representación analógica de una cantidad representa una parte de esa cantidad. Así, la representación aproximada de, por ejemplo, 12, contiene la representación de 8. Este tipo de representaciones también respeta las relaciones espaciales. En este sentido, estas representaciones guardan relaciones isomórficas con aquello que representan. De este modo, estas representaciones exhiben las características de las representaciones a las que, siguiendo a Barsalou y Prinz, llamo modales.

Carey (2009) señala que si bien las representaciones perceptivas y las de la cognición nuclear tienen el mismo vehículo, se diferencian por su grado de abstracción y por su participación en inferencias centrales. Ninguno de estos criterios parece adecuado para la distinción. Con respecto a la abstracción, ésta es una cuestión de grado y no está claro que una representación analógica de una magnitud, isomorfa con aquello que representa, sea más abstracta que, por ejemplo, las representaciones de profundidad o distancia. Con respecto al rol inferencial, Carey adopta una semántica de rol conceptual, la cual, como vimos en el capítulo 3, acarrea ciertas dificultades y, además, al igual que la abstracción, este criterio sólo permite establecer una distinción de grado. En consecuencia, considero que el criterio de control endógeno resulta más adecuado para distinguir entre las representaciones meramente perceptivas y las representaciones perceptivas conceptuales.

A lo largo del desarrollo, las representaciones de la cognición nuclear contribuyen a la adquisición de otras representaciones, en un vehículo distinto, de tipo amodal. Para estas categorías, no disponemos de principios innatos que guíen la recolección de datos. Estas teorías suelen asignar un rol fundamental al lenguaje en el desarrollo de los sistemas conceptuales que extienden el poder expresivo de los sistemas conceptuales nucleares, como ejemplificaré a continuación para el caso de los números.

Como sostiene Carey (2004, 2009), la historia de la adquisición de los números naturales involucra distintas etapas. A los dos años aproximadamente, los niños dominan una rutina de contar: en general, pueden enunciar una lista de números de manera ordenada. Sin

embargo, no dominan aun el uso de cada uno de los números, ni la relación que hay entre ellos como sucesivos. En esta primera etapa, se los puede describir como “conocedores de uno” [*one-knowers*]. A pesar de recitar la lista de números, no asignan el cardinal correcto a un conjunto. Por ejemplo, al pedirles un caramelo de un conjunto, entregan uno solo, pero al pedirles tres o cinco, en cambio, devuelven al azar cualquier cantidad mayor a uno, y no necesariamente una cantidad mayor para cinco que para tres (Carey 2004).

Etapas similares se observan entre los dos y los tres años y medio para el número dos y el número tres de modo que los niños pasan a ser “conocedores de dos” y “conocedores de tres”. Finalmente, aproximadamente a los tres años y medio, exhiben una habilidad de contar que involucra asignar la cardinalidad correcta a un conjunto. A esta edad, los niños pueden generar conjuntos con la cardinalidad de cualquier numeral de la lista que conocen. En esta etapa, al pedirles cuatro caramelos de un conjunto, tienden a devolver exactamente cuatro caramelos. Y, en caso de error, pueden recurrir a la rutina de contar para corregir la respuesta. Por ejemplo, si se les pide cuatro caramelos y devuelven cinco, ante la sugerencia de revisar si hay cuatro, los cuentan uno por uno y son capaces de sacar uno para arribar así a la solución correcta (Carey 2004, 2009). Según Carey, en esta etapa se observa que han adquirido la función sucesor: “si la palabra para el número x refiere a un conjunto con valor cardinal n , la siguiente palabra de la lista de números refiere a un conjunto con valor cardinal $n + 1$ ” (Carey 2004, p. 67).

Ahora bien, ¿cómo se explica el desarrollo a partir de las primeras habilidades numéricas al dominio de los conceptos de números naturales? Carey sostiene que la lista de números se aprende como una lista de ítems léxicos sin significado. Por eso en la etapa de saber sólo uno, o uno y dos, los niños pueden reproducir la lista de los números pero, sin embargo, no pueden identificar un conjunto como de cinco elementos. Esto implica que, frente a los conjuntos, continúan usando las habilidades preverbales (los sistemas de individuación de objetos y el de magnitudes analógicas). El lenguaje cumpliría un rol en la adquisición de los conceptos de números en dos sentidos. Primero, los recursos cuantificacionales propios del lenguaje, tales como la distinción singular-plural pueden contribuir al aprendizaje de los números. Carey sostiene que el primer significado de “número” que aprenden, *uno*, parte de los cuantificadores del lenguaje natural. La primera distinción de los “conocedores de uno” es entre uno y más de uno, se basa en la distinción morfológica entre singular y plural. La representación UNO podría ser un modelo en memoria a largo plazo de un conjunto de un único individuo construido a partir del sistema de individuación de objetos, más un procedimiento que determina que la palabra “uno” puede

ser aplicada a cualquier conjunto que se pueda poner en correspondencia con ese modelo. Lo mismo valdría para DOS y para TRES.

Aquí el lenguaje cumple una segunda función, al aprender la rutina de contar el niño aprende un orden entre los números. Sin embargo, el niño debe notar una analogía entre el orden de la lista de números y el orden en la serie de modelos, relacionados por la adición de un individuo. La disponibilidad de un sistema externo de símbolos ofrece un conjunto de marcadores [*placeholders*] que permite establecer analogías que involucran los distintos sistemas de cognición nuclear y realizar, en este caso, una inducción crucial y así, mediante procesos de *bootstrapping*, desarrollar representaciones conceptuales con un poder expresivo mayor que el de la cognición nuclear.

Las representaciones y los principios que conforman la cognición nuclear, sin embargo, no son reemplazados por los sistemas posteriores, sino que continúan operando en la vida adulta. Por ejemplo, en situaciones de presión temporal o en situaciones en las que no podemos contar, recurrimos al sistema de representación analógica de magnitudes si necesitamos comparar dos cantidades. El desempeño en estas tareas en tales circunstancias exhibe las marcas de los sistemas numéricos preverbales, tales como estar sujeto a la ley de Weber. Si, por ejemplo, en las cajas del supermercado, estimamos la cantidad de gente que hay en cada fila (o la cantidad de productos que hay en los changuitos de cada fila), tal estimación resulta más precisa cuanto mayor sea la diferencia en las cantidades. Estos efectos se han reportado en estudios controlados con estímulos de conjuntos de grandes cantidades de puntos (entre 20 y 200) en los que se debía determinar entre dos conjuntos cuál era mayor, para los cuales, a mayor razón entre los dos conjuntos, se observaba mayor precisión y menor tiempo de respuesta (Barth *et al.* 2006).

También hay evidencia de que una vez que se ha adquirido la lista de los numerales y se ha inducido la función sucesor, la lista de los numerales se integra con representaciones de magnitudes analógicas y, en consecuencia, éstas últimas intervienen en la resolución de ciertas tareas numéricas aun cuando se dispone de símbolos. Así, al igual que la estimación de la cantidad de elementos en grandes conjuntos, la comparación de números presentados simbólicamente, i.e. la determinación de cuál es más grande, también presenta la marca de la ley de Weber: los tiempos de respuesta son mayores al comparar 5 y 6 que al comparar, por ejemplo, 5 y 9 (Dehaene 1997).

Ahora bien, la apelación al lenguaje como un medio externo de símbolos que permite desarrollar nuevos sistemas representacionales podría sugerir que esta teoría propone una suerte de relativismo lingüístico, según el cual el lenguaje determina el sistema conceptual,

en tanto determina el modo en que recortamos el mundo (Whorf 1956). Sin embargo, esta no necesariamente es una consecuencia de esta propuesta. Primero, si se introdujera cierto relativismo, éste no sería radical como el que defiende Whorf, dado que se postula la continuidad conceptual de la cognición nuclear a lo largo de la vida, lo cual otorga una base conceptual compartible al margen del lenguaje. Segundo, si bien el lenguaje proporciona un sistema externo de símbolos que contribuyen a la adquisición de conceptos, estos símbolos no determinan el contenido de los conceptos que se adquieren, sino que proveen, en todo caso, una suerte de andamiaje para la construcción conceptual. Por último, si hay cierto grado de relativismo lingüístico o no, es una cuestión que debería determinarse empíricamente.

La razón por la que me detuve en el ejemplo de los números naturales, es que quiero sugerir que la adquisición de gran parte de nuestros conceptos podría tener una historia similar. Este tipo de arquitectura ofrece un vínculo interesante entre la percepción y la cognición y permite construir una explicación de la adquisición que toma en cuenta las habilidades tempranas, incluye cambios cualitativamente significativos a lo largo del desarrollo, pero también da cuenta del rol que los sistemas iniciales conservan en la vida adulta. Por otra parte, al igual que el pluralismo amplio y la arquitectura dual presentada en §2, esta teoría de la adquisición y desarrollo conceptual postula procesos conceptuales que operan sobre dos tipos de vehículos distintos, uno correspondiente a la cognición nuclear, que es modal en tanto corresponde a los *outputs* de sistemas analizadores de *input* (que serían similares a los sistemas de *input* dedicados tal como los caracteriza Prinz) y otro simbólico amodal. En la próxima sección retomaré los aspectos que tienen en común las propuestas duales de Sloman (1996, 2002) y Carey (2004, 2009), y señalaré las ventajas que ofrece adoptar propuestas de este tipo en conjunción con el pluralismo amplio.

6. El pluralismo amplio y las ventajas de una arquitectura dual

Este tipo de arquitectura resulta promisoria, por un lado, para explicar la adquisición de conceptos, como ilustra el caso de los números naturales presentado en §5. Como vimos en el capítulo 5, una de las ventajas principales de la teoría de proxitipos es su explicación de la adquisición de conceptos a partir del disparo de representaciones perceptivas. La propuesta de un sistema conceptual dual sugiere que disponemos de sistemas iniciales innatos que ante encuentros con los estímulos relevantes, generan representaciones de tipo perceptivo automáticamente, y que contribuyen en el desarrollo de sistemas simbólicos cualitativamente distintos y con un poder expresivo mayor al de los sistemas iniciales. Por otro lado, esta arquitectura parece adecuada, también, para explicar por qué algunas tareas activan estos

sistemas modales innatos, dando lugar a efectos tales como el costo en el cambio de modalidad en la verificación de propiedades, que vimos en el capítulo 5.

De modo análogo a lo que sostuve para los distintos tipos de estructuras semánticas y las tareas de categorización, el uso de representaciones perceptivas o simbólicas podría depender del tipo de tarea que se lleva a cabo o de las categorías particulares que se representan en cada caso. No todas las categorías estarán representadas en los dos sistemas. Como sostuve en el capítulo 5, el alcance de las representaciones modales está limitado a las categorías relativamente concretas, a cuyas instancias tenemos acceso mediante los sentidos. Así, los conceptos abstractos y teóricos exceden la capacidad representacional de los sistemas modales y están vehiculados por símbolos amodales.

Ahora bien, en los casos en los que una categoría está representada de modo redundante en los dos sistemas, podría suceder que ambos operaran en paralelo o que se activara uno de ellos en función de la tarea. Como sugieren los estudios relativos al razonamiento, que incluyen ejemplos como el de Linda citado en §3, el sistema modal resulta difícil de inhibir, requiriendo un esfuerzo y control atencional significativos. Sugiero que en algunos casos se “opta” por uno u otro sistema y en otros funcionan los dos sistemas en paralelo. Estos son los casos en los que se observa interferencia entre uno y otro. Como vimos en el capítulo 5, se han reportado casos en los que la carga en un sistema perceptivo interfiere con una tarea de verificación de propiedades conceptuales o la comprensión semántica de una oración interfiere con una instrucción motora.

Recordemos que en el caso de la carga en los sistemas perceptivos y la verificación de propiedades, la interferencia no se puede explicar apelando simplemente al uso de recursos generales como la memoria, puesto que la interferencia se observa solamente si aquello que se debe recordar y las propiedades que se deben verificar corresponden a la misma modalidad, pero no si corresponden a modalidades diferentes. De manera análoga, en el caso del efecto de compatibilidad acción-oración, éste sólo se observa cuando se produce una oposición entre la orientación o la dirección de aquello que una oración representa y la intención de ejecutar un movimiento. En este sentido, estos efectos resultan difíciles de explicar por medio de recursos generales, y parece tratarse de efectos dependientes de las operaciones efectuadas sobre representaciones modales.

El tipo de propuesta arquitectural que sugiero adoptar es compatible con la propuesta de múltiples tipos de estructuras semánticas. Las arquitecturas duales se basan en una distinción entre dos tipos de procesos o sistemas, dependiendo de los tipos de procesos y vehículos. Pero también postulan distintos subsistemas dentro de cada tipo, de modo que

pueden abarcar a las que Machery (2009) denomina teorías “multi-proceso” y no se comprometen con la tesis, quizás limitada dadas las características de nuestra mente, de que a cada capacidad cognitiva subyacen estrictamente dos procesos. Así, como señalé en §2, los tipos de procesos que operan sobre los prototipos y ejemplares corresponden al sistema modal, mientras que los que operan sobre definiciones, teorías y etiquetas atómicas, corresponden al sistema amodal.

En §2 sostuve que la adquisición de conceptos implica la apertura de un archivo correspondiente a una categoría, identificado con una etiqueta amodal. La propuesta de interacción entre los sistemas perceptivos, los conceptos modales y los archivos etiquetados requiere cierto desarrollo. Mi propuesta constituye una adaptación de la explicación de la adquisición que ofrece Fodor (2008), que vimos en el capítulo 4, y la que ofrece Barsalou (1999), que vimos en el capítulo 5. Como sostiene Barsalou, ante encuentros con propiedades en el mundo se activan representaciones en nuestros sistemas de *input* dedicados. Interpretando de modo amplio en qué consisten los sistemas de *input*, estos pueden incluir sistemas como el sistema de representación de magnitudes analógicas, que vimos en §5. Algunos de los *outputs* de estos sistemas dejan un registro que permite su reactivación posterior.

Los mecanismos encargados de almacenar los patrones de activación modales para su posterior reactivación en ausencia de los estímulos, también son los encargados de registrar correlaciones entre distintas representaciones y establecer asociaciones más o menos fuertes entre ellas en función del grado de correlación. En este sentido, son mecanismos de tipo asociativo similares a los que Sloman identifica con el sistema intuitivo. Estos mecanismos pueden dar lugar a la formación de prototipos y ejemplares, que serían conjuntos de rasgos modales. Hasta aquí la propuesta se basa en tesis convergentes de Barsalou (1999), Carey (2004, 2009), Sloman (1996) y Prinz (2002). Ahora bien, la explicación que propongo sugiere que cuando las representaciones perceptivas dejan un registro que pasa a estar bajo control endógeno para su posterior reactivación, ese registro “se archiva” bajo una etiqueta amodal. De este modo, las representaciones perceptivas pasarían a ser conceptuales cuando las etiquetamos. Esto no significa que el vehículo del pensamiento conceptual sea desde ese momento exclusivamente amodal, sino que las etiquetas amodales permiten individuar los registros almacenados de representaciones perceptivas *como* representaciones de una u otra categoría. Para pensar acerca de esa categoría podemos recurrir tanto a la etiqueta como a las representaciones modales (o multimodales, más precisamente).

Esta explicación resulta, a mi entender, compatible con parte de la propuesta fodoriana de disparo o adquisición causal-bruta de las etiquetas. Recordemos que Fodor reconoce que hay una primera etapa en la que se forma un prototipo que opera como mediador entre un concepto atómico y una propiedad en el mundo. Una vez que se aprende un prototipo, por medio de un proceso neurológico se adquiere un concepto atómico y la mente queda enlazada a una propiedad. Este proceso de adquisición, según Fodor, no se puede describir en términos de aprendizaje, en tanto no involucra ningún tipo de relación racional con la experiencia. Pues bien, este proceso neurológico podría interpretarse, precisamente, como el pasaje a control endógeno de determinados registros de activación de representaciones perceptivas. Sería una especie de “toma neurológica de control” de representaciones que, de otro modo, estarían bajo control de los estímulos.⁶⁰

Esta podría ser la explicación de la adquisición de conceptos concretos, en este sentido no es extensible a todos los tipos de conceptos que podemos tener. En otros casos, como vimos para el ejemplo de los números naturales, los procesos pueden ser más complejos y pueden involucrar procesos inductivos de mapeo y generalización. De este modo, mi propuesta ofrece una reinterpretación de la adquisición, tanto de los conceptos modales como de los amodales, basada en los procesos postulados por Barsalou (1999), Carey (2004, 2009) y Fodor (1998, 2008).

Ahora bien, la tesis de que los procesos conceptuales se distinguen principalmente en dos tipos dependiendo del vehículo de las representaciones sobre las que operan y de los principios que los rigen, puede suscitar diversas objeciones. En particular, la alusión a un sistema dual puede generar cierta resistencia puesto que, en muchos casos, en las arquitecturas duales se caracteriza uno de los sistemas como automático en el sentido de estar bajo control de los estímulos del entorno, mientras que el otro estaría bajo control interno del organismo.⁶¹ Si el sistema modal estuviera bajo control de los estímulos, estas representaciones no cumplirían con una de las características de las representaciones conceptuales que es estar bajo control endógeno.

En la presentación del tipo de arquitectura compatible con el pluralismo amplio propuse una caracterización que no implica que uno de los sistemas esté bajo control del entorno, sino que distingue los sistemas en función del tipo de vehículo que procesa cada uno

⁶⁰ Dado que Fodor afirma que puede haber distintos mecanismos de acceso semántico y que los prototipos son sólo uno de ellos, esta misma explicación valdría también para ejemplares.

⁶¹ Por ejemplo, Evans y Stanovich (2013) sostienen que una de las características principales de uno de los sistemas es ser automático, en tanto está bajo control de los estímulos, mientras que el otro puede “desacoplarse” y permite el pensamiento independiente del entorno.

y los principios que gobiernan tales procesos. Sin embargo, el vínculo entre las representaciones perceptivas y la respuesta automática de los sistemas perceptivos a los estímulos requiere ciertas aclaraciones. ¿Está el sistema modal bajo control endógeno? El control endógeno opera en dos sentidos. Por un lado, un sistema bajo control endógeno debe poder activarse en ausencia de los estímulos que hubieran ocasionado esas representaciones. Como vimos, al proponer que los vehículos conceptuales son modales, Barsalou y sus colegas afirman precisamente esto, que representaciones *del mismo tipo* que aquellas que se darían en la percepción de los estímulos, se pueden activar con independencia de tales estímulos para llevar a cabo diversos procesos cognitivos. La tesis afirma que son representaciones vehiculadas en el mismo tipo de formato, no que son exactamente las mismas representaciones.

Si bien los sentidos en tanto sistemas de *input* dedicados están bajo control del entorno y no se pueden “apagar” por control del organismo, los registros que se almacenan para poder reactivar las representaciones perceptivas en ocasiones futuras no necesariamente comparten esta característica. Los simuladores (i.e. los registros de patrones de activación de representaciones perceptivas) son más complejos y las simulaciones están bajo control endógeno. Un simulador particular es el resultado de un proceso de abstracción basado en experiencias perceptivas particulares y, en tanto tal, no está bajo control del entorno. No todo lo que representan los sistemas perceptivos en una situación dada se registra y almacena como parte de la misma categoría. Por ejemplo, imaginemos que una de las primeras experiencias perceptivas que una persona tiene tomando mate sucede en el jardín de una casa, una tarde de verano, acompañado de bizcochitos de grasa. Las características asociadas a MATE probablemente incluyan la propiedad de ser amargo, características de la yerba y de la bombilla. Pero no se asocian a las representaciones de esta categoría las características del jardín, ni del verano, ni tantas otras que se perciben simultáneamente. En ocasiones posteriores, esa persona podrá pensar en MATE sin pensar en jardines.

De este modo, en los sistemas propuestos por los defensores de los vehículos modales, no todas las representaciones perceptivas son conceptuales, sino que hay una distinción entre los estados meramente perceptivos y los estados perceptivos que son conceptuales. Lo que no hay es una diferencia de vehículo. Las representaciones perceptivas en tanto conceptos son distintas de las representaciones perceptivas en general, porque están organizadas como representaciones de categorías y son independientes del entorno. Así, las representaciones conceptuales modales pueden estar bajo control del organismo, en tanto implican cierta abstracción respecto de las representaciones que las ocasionaron y se pueden activar en

ausencia de los estímulos. Ahora bien, ¿se pueden inhibir estas representaciones en presencia de los estímulos? Si bien los sistemas perceptivos en sí no se pueden inhibir por control endógeno ante los estímulos correspondientes, sí se podría inhibir, por lo menos en principio, y en cierta medida, la activación de la representación multimodal a la que están asociadas en la memoria a largo plazo. En este sentido, la representación modal conceptual no está bajo control del entorno.

Puede haber casos particulares en los que inhibir la activación de las simulaciones sea más difícil. Esto, en lugar de ser una desventaja, puede implicar por el contrario una ventaja explicativa, dado que permite dar cuenta de casos en los que, por ejemplo, no podemos evitar reconocer nuestro nombre. Una analogía con el lenguaje podría servir para ilustrar esto. El sistema de procesamiento de lenguaje está en cierta medida bajo control de los estímulos: no podemos evitar procesar como lenguaje los estímulos lingüísticos que percibimos. Sin embargo, en ocasiones, podemos distraer nuestra atención de ellos y convertirlos en “ruido de trasfondo”, por ejemplo si estamos intentando leer en un bar o en un colectivo. En estas situaciones, el sistema no está “apagado” o desconectado de los estímulos, pero nuestra atención está desviada. Una situación muy común es que, estando en este estado en el que no entendemos lo que se dice a nuestro alrededor, si alguien pronuncia nuestro nombre, automáticamente se interrumpe el estado de concentración y lo reconocemos, i.e. se destaca esa palabra entre el sonido que hasta ese momento era percibido como ruido de fondo y no como lenguaje.

Sugiero que podría suceder algo similar con las representaciones conceptuales modales. De la misma manera que hay una interacción entre los sistemas conceptuales modal y amodal, como se manifiesta en los casos de conflicto y competencia por los recursos, hay también una interacción entre los sistemas modales (i.e. puramente perceptivos o no conceptuales) y el sistema conceptual modal. Si bien las representaciones conceptuales modales están bajo control endógeno en los dos sentidos mencionados, i.e. se pueden activar en ausencia del estímulo y pueden inhibirse en presencia del estímulo, el vínculo estrecho que mantienen con las representaciones perceptivas que están efectivamente bajo control de los estímulos (y que no se pueden inhibir por voluntad del sujeto) produce que en algunos casos la inhibición resulte difícil.

Por otra parte, cabe mencionar que la distinción que se establece en las teorías de procesos duales entre procesos automáticos y procesos controlados, no necesariamente coincide con la distinción entre los sistemas dependientes o independientes de los estímulos, sino que en algunos modelos parece depender del control voluntario que puede ejercer el

sujeto sobre los procesos. Por ejemplo Sloman (2002) señala que respecto de los procesos intuitivos, sólo nos percatamos del *output* o resultado, mientras que respecto de los procesos deliberativos nos percatamos, además, del proceso en sí (i.e. de las reglas que gobiernan el razonamiento y de los pasos mediante los cuales arribamos a la respuesta). De este modo, la propuesta parece poder hacer frente a estas posibles objeciones.

Por último, se puede señalar como ventaja de la adopción de una arquitectura dual, el hecho de que este tipo de arquitectura se postula para explicar fenómenos relativos a capacidades cognitivas diversas, como mencioné en §3, muchos de los cuales, probablemente, involucran habilidades conceptuales. Así, articular una arquitectura de este tipo con una teoría de conceptos podría facilitar la integración entre explicaciones de distintos aspectos de la cognición. Si bien cada teoría de sistemas o procesos duales implica compromisos específicos y modelos particulares, la propuesta de un sistema conceptual dual podría integrarse con estos modelos. De hecho, este es un objetivo interesante como programa de investigación. Si bien no se ha establecido aun un marco unificado que delimite una arquitectura cognitiva común, capaz de abarcar todas las propuestas de sistemas duales, un buen punto de partida para desarrollar este marco podría ser aquel que privilegie la integración de las distintas explicaciones en función de los sistemas representacionales sobre los que operan. En este sentido, una teoría de las representaciones mentales conceptuales formará parte de este tipo de propuesta.

Conclusiones

Los debates entre los enfoques principales respecto de la estructura semántica y el vehículo de los conceptos mantenidos aproximadamente entre la década de 1970 y hasta hace algunos años, parecen arrojar el siguiente diagnóstico para las teorías de conceptos: no hay un único tipo de estructura semántica capaz de satisfacer los requisitos explicativos de una teoría adecuada de conceptos. Aquellos debates adoptaban como supuesto el principio de uniformidad, según el cual la clase de los conceptos es homogénea. Dadas las dificultades para dar cuenta de los requisitos explicativos por medio de un único tipo de estructura semántica, las recientes propuestas híbridas, pluralistas y eliminativistas postulan que la clase de los conceptos constituye una clase heterogénea, que incluye estructuras semánticas de distinto tipo. El objetivo de esta tesis fue defender una propuesta pluralista que hiciera extensiva esta tesis a los vehículos conceptuales. Si bien la propuesta de que los conceptos pueden tener distintos tipos de estructuras semánticas ha ganado consenso en los últimos años, con respecto a los vehículos conceptuales suele adoptarse el principio de la uniformidad.⁶²

Así, desarrollé la propuesta del *pluralismo amplio*, que afirma que los conceptos pueden tener tanto distintos tipos de estructura semántica interna, como distintos tipos de vehículos. Sostuve que los conceptos pueden estar estructurados como definiciones, prototipos, ejemplares o teorías, o pueden ser atómicos, y que éstas estructuras pueden estar vehiculadas en un formato amodal o en un formato modal, y sugerí qué combinaciones de estructura interna y formato parecen más adecuadas, en función de una propuesta de arquitectura cognitiva dual.

La estrategia general de defensa del pluralismo amplio puede leerse como una extensa inferencia a la mejor explicación, cuyas premisas establecí y defendí en los capítulos 2 a 5. Analicé los enfoques principales respecto de la estructura semántica y los vehículos conceptuales: los enfoques de definiciones, prototipos, ejemplares, teorías y etiquetas atómicas, y los de vehículos amodales y modales. El objetivo al analizar cada uno de los enfoques no fue retroceder al principio de uniformidad, sino fundamentar por qué resulta ventajoso para el pluralismo amplio postular cada una de estas estructuras semánticas y vehículos. Por este motivo, al reconstruir sus tesis y ventajas principales, dediqué especial atención a la evidencia a su favor. El pluralismo amplio pretende ofrecer un enfoque empírico

⁶² Se puede mencionar como excepciones a Dove (2009) y Carey (2009).

acerca de nuestras representaciones mentales conceptuales, aunque esquemático y preliminar. En este sentido, pretende dar cuenta de los resultados de los últimos años de investigación psicológica acerca de cómo representamos el mundo.

Los argumentos basados en las dificultades de determinados enfoques para satisfacer los requisitos explicativos pueden despertar una objeción relativa a si resulta legítimo o adecuado evaluar todos los enfoques en función de los mismos requisitos, dado que no hay un consenso total respecto de los fenómenos que deben explicar las teorías de conceptos y cada enfoque prioriza distintos subconjuntos de requisitos. Para justificar este tipo de estrategia, analicé los principales requisitos explicativos que figuran en la literatura y propuse seleccionar entre ellos los requisitos de contenido, incluyendo en este requisito el de contenido cognitivo (que permite explicar los casos de diferencias entre conceptos coextensivos y semejanzas entre conceptos con distinta referencia), alcance, composicionalidad, categorización, adquisición y publicidad. Esta selección se fundamenta en la idea de que las teorías de conceptos se proponen para dar cuenta tanto de cuestiones metafísicas de nuestro pensamiento, relativas a su capacidad representacional y poder expresivo (y por eso deben explicar el contenido, el alcance y la composicionalidad), como al uso de los conceptos en procesos cognitivos (y por eso deben explicar la adquisición, la categorización y la publicidad).

Como parte de la defensa de estos requisitos discutí la distinción que establecen algunos autores entre teorías filosóficas (que serían las que se concentran en los requisitos relacionados con el poder expresivo del pensamiento) y las psicológicas (que se ocuparían de los requisitos relativos al uso de conceptos en procesos cognitivos) (Machery 2009, Fodor 1999, 2000b). Contra esta distinción sostuve que no refleja adecuadamente los objetivos explicativos de gran parte de la investigación de conceptos, la cual integra la mayoría de los requisitos propuestos en la literatura. Asimismo, esta distinción parece ignorar que los debates entre los defensores de las denominadas teorías psicológicas y filosóficas han contribuido al avance de la investigación, en lugar de delimitarla. Por último, los distintos requisitos están relacionados entre sí, de modo que el abordaje de algunos de ellos puede implicar compromisos respecto del resto.

La selección del conjunto de requisitos fundamentales mencionado me permitió evaluar los principales enfoques respecto de la estructura semántica de los conceptos, así como respecto de sus vehículos. De esta manera, sostuve que las definiciones pueden resultar adecuadas para explicar el contenido y la composicionalidad de los conceptos, pero su alcance es sumamente limitado. Si bien podrían ser adecuadas para algunos tipos de

conceptos, como los lógicos y los matemáticos, la mayoría de los conceptos léxicos no muestra estructura definicional en los experimentos psicológicos.

Los enfoques de prototipos tienen la ventaja de explicar satisfactoriamente la categorización a partir de una función de similitud que compara rasgos percibidos con los rasgos representados en cada concepto y, por otra parte, la ventaja de explicar la adquisición, como el registro y la abstracción de regularidades estadísticas a partir de la experiencia. Asimismo, determinados efectos observados en tareas de categorización se explican fácilmente apelando a prototipos. Estos son los efectos de tipicidad: el hecho de que las instancias típicas de las categorías se categorizan más rápido y con menor tasa de error que las atípicas, y resulta más veloz el aprendizaje de la pertenencia a una categoría para las instancias típicas que para las atípicas. Sin embargo, este enfoque no puede explicar otros efectos asociados a la categorización: los efectos de ejemplares (relacionados con la familiaridad con determinadas instancias) y los efectos causales (relativos a la asimetría en la relevancia diagnóstica de los rasgos en función de su estatus causal). Ahora bien, la dificultad principal de los prototipos, radica en que no se combinan composicionalmente. Al combinar dos prototipos, suelen perderse rasgos de los conceptos que se combinan o surgir propiedades emergentes. Estos efectos resultan difíciles de explicar por medio de reglas de combinación y sin apelar a estructuras conceptuales de otro tipo, como los ejemplares o las teorías. Por otra parte, su alcance tiene ciertas limitaciones, puesto que hay conceptos que no tienen prototipo, en particular los formados por operaciones booleanas, como NO GATO.

El enfoque de ejemplares tiene la ventaja de explicar los efectos de ejemplares observados en la categorización. En particular, predice y explica los efectos que tiene la familiaridad con determinadas instancias de una categoría en los tiempos de reacción para categorizar instancias similares a esos ejemplares. Contra lo que predicen las teorías de prototipos, muchos estudios indican que podemos categorizar instancias atípicas de una categoría igualmente o más rápido que instancias típicas, en caso de haber tenido experiencias previas con instancias particulares similares al ejemplar atípico en cuestión. Por otro lado, da cuenta del rol que cumplen nuestras interacciones con instancias particulares de categorías en la adquisición de conceptos. Sin embargo, este enfoque enfrenta una serie de dificultades. No consigue dar cuenta del alcance, dado que no se puede explicar la posesión de conceptos teóricos, e.g. ELECTRÓN, como un conjunto de ejemplares. Tampoco explica la composicionalidad, dado que no queda claro cómo podrían combinarse distintos conjuntos de ejemplares para formar un concepto nuevo. Por último, no explica la publicidad, puesto que no resulta plausible que distintos individuos representen los mismos ejemplares, sino que

dado que cada uno tiene experiencias con distintos ejemplares, cada uno tendrá conceptos distintos.

El enfoque de teoría-teoría ofrece ventajas en la explicación de los efectos causales en la categorización, el alcance y la adquisición de conceptos. Por ejemplo, da cuenta de la asimetría en tareas de categorización entre la relevancia de los rasgos que representan causas y aquellos que representan efectos. Sin embargo, tiene limitaciones para explicar el contenido. Este es un requisito del cual generalmente este enfoque no se ocupa explícitamente. Sin embargo, cualquier teoría que asuma que los conceptos son representaciones mentales de categorías, acarrea compromisos semánticos, de modo que resulta pertinente explicitar estos compromisos y, partiendo de ellos, evaluar si se podría dar cuenta de este requisito. El principal problema en relación a este requisito es que para muchos conceptos disponemos únicamente de teorías esquemáticas e incompletas que no resultan suficientes para determinar a qué categoría refieren. Como consecuencia, debe haber un modo de individuar el contenido de los conceptos, más allá de su estructura teórica.

Sostuve, asimismo, que todos los enfoques que postulan que los conceptos poseen una estructura semántica interna comparten el desafío de explicar la publicidad. Estos enfoques postulan que el contenido de los conceptos está estructurado, ya sea por relaciones mereológicas con los rasgos que conforman su estructura semántica o por relaciones inferenciales en la red de creencias de la que participan, dependiendo de cada teoría en particular. Como consecuencia, el contenido de los conceptos depende de sus relaciones con otros conceptos. Por esta razón, se ven amenazados por la posibilidad de que el sistema conceptual resulte holista, i.e. que el contenido de un concepto dependa de sus relaciones con el sistema completo de interrelaciones conceptuales. Si esto fuera así, no podrían dar cuenta de la publicidad, dado que no resulta plausible que distintos individuos compartan sistemas completos de relaciones entre conceptos.

Con respecto al atomismo conceptual, sostuve que resulta especialmente adecuado para dar cuenta de los requisitos de alcance, composicionalidad y publicidad y, si bien la semántica que propone resulta adecuada para explicar el contenido intencional (y es, de hecho, la que propuse adoptar en el pluralismo amplio), tiene sin embargo dificultades para explicar la individuación de conceptos coextensivos. Por otra parte, el atomismo no ofrece una explicación completa de la adquisición, dado que no explica cómo se vinculan los conceptos atómicos y los mecanismos de acceso semántico que son intermediarios entre los conceptos y el mundo. Respecto de la categorización exhibe un problema similar.

De este modo, con respecto a la estructura conceptual interna, sostuve que todas las

propuestas tienen dificultades para explicar algunos de los requisitos fundamentales, pero cada una resulta adecuada para explicar algunos de ellos. En conjunto, estas entidades permitirían dar cuenta de todos los requisitos fundamentales. De modo que sostuve que la mejor explicación de los requisitos fundamentales involucra postular que la clase de los conceptos incluye definiciones, prototipos, conjuntos de ejemplares, teorías y etiquetas atómicas.

Señalé que un ejemplo de las ventajas de postular distintos tipos de estructuras semánticas está dado por la teoría de proxitipos. Sin embargo, sostuve que esta teoría presenta numerosas desventajas como consecuencia de postular que todos los conceptos están vehiculados por representaciones perceptivas. Sostuve que la adopción de representaciones perceptivas permite explicar la adquisición de conceptos, así como dar cuenta de una serie de efectos empíricos relativos a la interacción entre los mecanismos perceptivos y conceptuales en tareas de categorización. Sin embargo, mostré que la teoría de proxitipos no consigue dar cuenta de la individuación del contenido de los conceptos, de la representación de conceptos teóricos y abstractos, ni de la composicionalidad recurriendo únicamente a representaciones perceptivas.

Con respecto a los vehículos conceptuales, señalé que el LDP no presenta limitaciones en relación a los distintos requisitos, a diferencia de las representaciones perceptivas que tienen dificultades con los requisitos de contenido, alcance y composicionalidad. Sin embargo, sostuve que la evidencia a favor de las representaciones perceptivas resulta difícil de acomodar por los enfoques amodales. De este modo, sugerí que la adopción de los dos tipos de vehículos es la mejor manera de dar cuenta de los requisitos fundamentales.

Así, propuse una versión de propuesta pluralista, el pluralismo amplio, que postula tanto distintos tipos de estructura semántica como distintos tipos de vehículos. Esta propuesta se diferencia de las teorías contemporáneas híbridas y pluralistas en postular una multiplicidad de vehículos conceptuales. Comparte con ellas el compromiso con dos tesis centrales: la clase de los conceptos incluye estructuras semánticas heterogéneas correferenciales que pueden activarse de modo independiente y estas representaciones están asociadas en la memoria a largo plazo como representaciones de la misma categoría. En este sentido, ofrecí una reconstrucción novedosa del debate entre los enfoques híbridos y pluralistas contemporáneos, que vincula la distinción entre ambos a la discusión con el eliminativismo.

En su defensa del eliminativismo, Machery distingue entre teorías híbridas y

pluralistas con el objetivo de establecer la preferencia de las teorías pluralistas por sobre las híbridas. Una vez establecida esta preferencia, a partir de la heterogeneidad de estructuras conceptuales que postula el pluralismo, Machery sostiene que, dada tal heterogeneidad, los conceptos no constituyen una clase natural y, por lo tanto, la noción de “concepto” no resulta de utilidad para la ciencia. Una estrategia para resistir el eliminativismo consiste en mostrar que las estructuras conceptuales son en algún sentido homogéneas, de modo que constituyen una clase de interés científico. Postular que todos los conceptos tienen una estructura híbrida común permitiría así rechazar el eliminativismo. Sin embargo, pueden adoptarse otras estrategias para bloquear la conclusión eliminativista. Sostuve, siguiendo a Weiskopf (2009a), que los conceptos constituyen una clase funcional, delimitada a partir de los roles que satisfacen los conceptos en nuestra vida mental. Esta clase puede resultar de interés científico, dado que pueden establecerse generalizaciones fructíferas en relación a los procesos de, e.g. combinación conceptual, y almacenamiento y recuperación de conceptos. De este modo, puede defenderse que los conceptos constituyen una clase de interés científico sin apelar a una homogeneidad casi artificial entre sus instancias.

La reconstrucción de estos debates me permitió, asimismo, elucidar las dos tesis centrales que a mi entender comparten los enfoques híbridos y pluralistas, y a las que adscribe el pluralismo amplio. Estas tesis revelan cierta tensión entre dos niveles de individuación de los conceptos. Por un lado, hay un nivel que denominé arquitectural, relativo a la independencia que mantienen las distintas estructuras correferenciales entre sí, por la cual cada una puede ser activada en la memoria de trabajo independientemente del resto. Por otro lado, hay un nivel semántico, en el cual distintas estructuras se asocian como correferenciales.

La propuesta del pluralismo amplio intenta ofrecer una respuesta a esta tensión, apelando, por un lado, a la metáfora de los archivos mentales, según la cual nuestro sistema conceptual estaría organizado en archivos, cada uno correspondiente a una categoría. Los distintos conceptos estarían agrupados o asociados en virtud de referir a la misma categoría (aunque esta referencia no sea del mismo tipo). Sugerí individuar conceptos en virtud de su estructura semántica interna o contenido cognitivo, i.e. por cómo representan aquello que representan. Así, esta propuesta reconoce la independencia de los distintos tipos de representaciones que pueden activarse independientemente en la memoria de trabajo. A la vez, da cuenta de la asociación de los distintos conceptos correferenciales, puesto que estos forman parte del mismo archivo. Sugerí que los archivos se individuarían en función de la propiedad que representan.

El pluralismo amplio parece adecuado para dar cuenta de los requisitos fundamentales. En relación a la *categorización*, la inclusión de prototipos, ejemplares y teorías permitiría dar cuenta de los distintos efectos observados, que dependerían del uso de uno u otro tipo de estructura conceptual. Con respecto a la determinación de qué tipo de estructura interviene en cada tarea, sugerí que se trata de una cuestión de dependencia contextual y que la activación de una u otra estructura depende tanto de las condiciones específicas de la tarea en cuestión (i.e. el modo en que se presentan los estímulos y el tipo de respuesta que demanda la tarea) como de la categoría en cuestión.

Con respecto al *contenido*, la apelación a una organización en archivos mentales permitiría articular los distintos modos de referir que tiene cada una de las estructuras semánticas postuladas. Cada estructura semántica presenta un contenido cognitivo, dado por su estructura interna. A la vez, las distintas representaciones correferenciales estarían asociadas a un mismo archivo, individuado por su relación con el mundo, i.e. por la categoría que representa. La asociación de distintas representaciones a un mismo archivo, a su vez, depende de cuestiones epistémicas relativas a qué representaciones consideramos correferenciales. La *composicionalidad* estaría garantizada por la inclusión de símbolos atómicos, mientras que los efectos asociados a la combinación de conceptos como el de propiedades emergentes o pérdida de propiedades de los conceptos iniciales pueden explicarse por medio del resto de las estructuras semánticas. En este sentido sugerí que un modelo de combinación como el RCA que propone Prinz resultaría adecuado. El pluralismo amplio puede explicar el *alcance*, dado que postula tanto distintos tipos de estructuras semánticas como distintos tipos de vehículos. Las representaciones perceptivas ofrecerían un nexo entre percepción y cognición, facilitando la explicación de la adquisición de conceptos ante encuentros con sus instancias.

Dadas las tesis del pluralismo amplio, éste enfrenta posibles dificultades en relación al requisito de *publicidad*. Primero, dado que incorpora estructuras semánticas que es plausible que no sean compartidas (como los conjuntos de ejemplares), habría por lo menos algunas estructuras que distintos individuos no comparten. Sin embargo, puesto que también postula que cada archivo está identificado por una etiqueta atómica, y éstas son compartibles, compartiríamos en ese sentido representaciones de categorías. Segundo, dado que postula representaciones con una estructura semántica interna, esta propuesta se ve amenazada por el holismo. Sin embargo, sugerí que se puede adoptar la estrategia localista, de modo de sostener que el conjunto de relaciones entre conceptos que resultan constitutivas del contenido es limitado.

Para explicar la articulación entre los distintos tipos de representaciones conceptuales y los modos de procesamiento que cada una conlleva, sostuve que las tesis del pluralismo amplio resultan compatibles con una arquitectura dual como la que propone Sloman (1996, 2014). Así, habría dos sistemas conceptuales, uno asociativo, que opera sobre representaciones perceptivas, y cuyas operaciones reflejan similitud y estructura espacio-temporal; y otro simbólico, que opera sobre representaciones amodales a partir de reglas formales sensibles a la estructura sintáctica. Sugerí a qué sistema corresponde cada una de las estructuras conceptuales postuladas. Sostuve que las etiquetas atómicas, las teorías y las definiciones corresponden al sistema deliberativo simbólico, mientras que los prototipos y los conjuntos de ejemplares corresponden al sistema asociativo modal. Por último, sugerí adoptar la distinción que propone Carey (2009) entre cognición nuclear y cognición en general, para dar cuenta de la *adquisición* de conceptos y explicar cómo intervienen en la misma los distintos tipos de estructuras conceptuales que postula el pluralismo amplio.

De este modo, la propuesta de pluralismo amplio no sólo permitiría integrar las ventajas de distintos enfoques de conceptos, sino que también ofrece una visión convergente con teorías que provienen de otros ámbitos de la ciencia cognitiva. Considero que una de las ventajas del pluralismo amplio es la posibilidad de integrar explicaciones de distintas capacidades en un marco teórico unificado. Puede parecer casi paradójico el énfasis en la unificación planteado desde una defensa del pluralismo. Sin embargo, la atención a las características específicas de cada tipo de estructura, y el reconocimiento de que cada una resulta más adecuada para explicar distintos tipos de efectos, no implica que éstas sean estructuras en sistemas aislados sin interacción o que no colaboren entre sí para llevar a cabo determinadas tareas. La posibilidad de abarcar distintas estructuras en una teoría integradora es una ventaja y no debería conllevar una pérdida de especificidad.

Así, la propuesta del pluralismo amplio constituye una sugerencia esquemática de una línea de investigación que parece fructífera. En este sentido, la elaboración y puesta a prueba de modelos específicos de articulación de los distintos tipos de estructuras semánticas y vehículos postulados permitirá precisar las características específicas de los distintos tipos de estructuras conceptuales, así como de su interacción. En lugar de seguir buscando un tipo de entidad ideal que permita explicar los requisitos fundamentales de una teoría de conceptos, ofrecer un enfoque a partir de las estructuras y los vehículos que cuentan con apoyo empírico parece una empresa promisoría.

Referencias

- Ahn, W. (1998). Why are different features central for natural kinds and artifacts?: the role of causal status in determining feature centrality. *Cognition*, 69, 135-178.
- Ahn, W., & Luhmann, C. (2004). Demystifying theory-based categorization. En L. Gershkoff-Stowe, & D. Rakison (Eds.), *Building Object Categories in Developmental Time* (pp. 277-300). Mahwah, NJ: Lawrence Erlbaum Associates.
- Ahn, W., Gelman, S., Amsterlaw, J., Hohenstein, J., & Kalish, C. (2000). Causal status effect in children's categorization. *Cognition*, 76, B35-B43.
- Anderson, J., & Betz, J. (2001). A hybrid model of categorization. *Psychonomic Bulletin & Review*, 8(4), 629-647.
- Apperly, I. A., & Butterfill, S. A. (2009). Do Humans Have Two Systems to Track Beliefs and Belief-Like States? *Psychological Review*, 116, 953-970.
- Ashby, F. G., & Ell, S. (2002). Single versus multiple systems of learning and memory. En J. Wixted, & H. Pashler (Eds.), *Steven's handbook of experimental psychology* (3 ed., Vol. 4, pp. 655-692). New York: Wiley.
- Barberis, S. (2008). Two arguments against proxytype compositionality. *Presentado en I Workshop on Philosophy and Cognitive Science: The Philosophy of Psychology of Jesse J. Prinz / III Encuentro sobre Lenguaje*. Buenos Aires.
- Barsalou, L. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 11(4), 629-654.
- Barsalou, L. (1987). The Instability of Graded Structure: Implications for the Nature of Concepts. En NeisserUlric (Ed.), *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. Cambridge: Cambridge University Press.
- Barsalou, L. (1999). Perceptual Symbol Systems. *Behavioral & Brain Sciences*, 22, 577-660.
- Barsalou, L., Pecher, D., Zeelenberg, R., Simmons, W. K., & Hamann, S. (2005). Multimodal simulation in conceptual processing. En W. Ahn, R. Goldstone, B. Love, A. Markman, & P. Wolff (Eds.), *Categorization inside and outside the lab: Essays in honor of Douglas L. Medin* (pp. 249-270). Washington, DC: American Psychological Association.
- Barsalou, L., Simmons, W. K., Barbey, A., & Wilson, C. (2003). Grounding conceptual knowledge in modality-specific systems. *TRENDS in Cognitive Science*, 7, 84-91.

- Barsalou, L., Solomon, K., & Wu, L.-L. (1999). Perceptual simulation in conceptual tasks. En M. Hiraga, C. Sinha, & S. Wilcox (Eds.), *Cultural, psychological and typological issues in cognitive linguistics* (pp. 209-228). Amsterdam: John Benjamins Publishing Company.
- Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition*, 98, 199-222.
- Block, N. (1986). Advertisement for a Semantics for Psychology. En P. French, T. Uehling, & H. Wettstein (Eds.), *Midwest Studies in Philosophy, X* (pp. 615-678). Minneapolis: University of Minnesota Press.
- Borges, J. L. (1974). Funes el memorioso. En *Obras completas 1923-1972*. Buenos Aires: Emecé Editores.
- Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies*, 127-148.
- Braisby, N. (2005). Similarity and categorisation: getting dissociations in perspective. En K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Cognitive Science Society* (pp. 150-155). Chicago: Erlbaum.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, Mass.: MIT Press.
- Carey, S. (1988). Conceptual differences between children and adults. *Mind and Language*, 3, 167-181.
- Carey, S. (1992). The origin and evolution of everyday concepts. In R. Giere (Ed.), *Cognitive Models of Science* (pp. 89-128). Minneapolis: University of Minnesota Press.
- Carey, S. (1999). Knowledge Acquisition: Enrichment or Conceptual Change? En E. Margolis, & S. Laurence (Eds.), *Concepts: Core readings* (pp. 459-487). Cambridge, MA: MIT Press.
- Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus*, 59-68.
- Carey, S. (2009). *The origin of concepts*. Oxford: Oxford University Press.
- Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. En L. Hirschfeld, & S. Gelman (Eds.), *Mapping the mind: domain specificity in cognition and culture* (pp. 169-200). New York: Cambridge University Press
- Carey, S., & Spelke, E. (1996). Science and core knowledge. *Philosophy of science*, 63, 515-533.
- Connolly, A., Fodor, J., Gleitman, L., & Gleitman, H. (2007). Why stereotypes don't even make good defaults. *Cognition*, 103, 1-22.

- Costello, F., & Keane, M. (2000). Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science*, 24(2), 299-349.
- Cushman, F., & Grene, J. (2012). Finding faults: How moral dilemmas illuminate cognitive structure. *Social Neuroscience*, 7(3), 269-279.
- Cushman, F., Young, L., & Hauser, M. (2006). The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm. *Psychological Science*, 17, 1082-1089.
- De Rosa, R. (2005). Prinz's Problematic Proxytypes. *The Philosophical Quarterly*, 55 (221), 594-606.
- Dehaene, S. (1997). *The number sense*. New York: Oxford University Press
- Destéfano, M. (2009). Los límites de la estrategia lingüística de Jesse Prinz. *Presentado como parte de la mesa "Conceptos: Un problema para el empirismo", VI Jornadas de Filosofía Teórica*. UNC, Córdoba.
- Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, 22, 735-808.
- Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition*, 110, 412-431.
- Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.
- Edwards, K. (2010). Unity Amidst Heterogeneity in Theories of Concepts. *Behavioral and Brain Sciences*, 33, 210-211.
- Evans, G. (1982). *The varieties of reference*. Oxford: Oxford University Press.
- Evans, J. (2003). In two minds: dual-process accounts of reasoning. *TRENDS in Cognitive Sciences*, 17, 454-459.
- Evans, J., & Stanovich, K. (2013). Dual Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, 8(3), 223-241.
- Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: evidence from infant's manual search. *Developmental Science*, 6, 568-584
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *TRENDS in Cognitive Science*, 8, 307-314.
- Fischer, S. (1916). The process of generalizing abstraction; and its product, the general concept. *Psychological Monographs*, 21, 1-209.
- Fodor, J. (1975). *The language of thought*. New York: Thomas Y. Crowell Company.

- Fodor, J. (1981). The present status of the innateness controversy. In J. Fodor, *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT Press.
- Fodor, J. (1983). *The modularity of mind*. Cambridge, Mass.: MIT Press.
- Fodor, J. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.
- Fodor, J. (1990). *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.
- Fodor, J. (1994). *The Elm and the Expert: Mentalese and its semantics*. Cambridge, MA: MIT Press.
- Fodor, J. (1999). *Conceptos*. Barcelona: Gedisa.
- Fodor, J. (2000a). *The mind doesn't work that way*. Cambridge, MA: MIT Press.
- Fodor, J. (2000b). Replies to Critics. *Mind & Language*, 15(2 & 3), 350–374.
- Fodor, J. (2001). Doing without what's within: Fiona Cowie's critique of nativism. *Mind*, 110, 99-148.
- Fodor, J. (2008). *LOT 2*. Oxford: Clarendon Press.
- Fodor, J., & Lepore, E. (1992). *Holism: A shopper's guide*. Oxford: Blackwell
- Fodor, J., & Lepore, E. (1996). The red herring and the pet fish: why concepts still can't be prototypes. *Cognition*, 58, 253-270.
- Fodor, J., & Pylyshyn, Z. (1995). Connectionism and cognitive architecture: A critical analysis. En C. Macdonald, & G. Macdonald (Eds.), *Connectionism: Debates on psychological explanation* (pp. 90-163). Oxford: Blackwell.
- Frege, G. (1969). Sobre el sentido y la denotación. In T. Moro Simpson (Ed.), *Semántica Filosófica: problemas y discusiones*. Buenos Aires: Siglo XXI.
- Gelman, S. (2004). Psychological essentialism in children. *TRENDS in Cognitive Sciences*, 8, 404-409.
- Glenberg, A., & Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558-565.
- Gonnerman, C., & Weinberg, J. (2010). Two uneliminated uses for “concepts”: Hybrids and guides for inquiry. *Behavioral and Brain Sciences*, 33, 211-212.
- Gopnik, A. (2011). The theory theory 2.0: probabilistic models and cognitive development. *Child Development Perspectives*, 5, 161-163.
- Gopnik, A., & Meltzoff, A. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.
- Gopnik, A., & Schulz, L. (2004). Mechanisms of theory formation in young children. *TRENDS in Cognitive Sciences*, 8, 371-377.

- Gopnik, A., & Wellman, H. (2012). Reconstructing constructivism: causal models, bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*, 1085-1108.
- Greene, J. (2009). Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Experimental Social Psychology*, *45*(3), 581-584.
- Greene, J., Nystrom, L., Engell, A., Darley, J., & Cohen, J. (2004). The Neural Bases of Cognitive Conflict and Control in Moral Judgment. *Neuron*, *44*, 389-400.
- Hampton, J. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *18*, 441-461.
- Hampton, J. (1995). Testing the prototype theory of concepts. *Journal of Memory and Language*, *34*, 686-708.
- Hampton, J. (1997). Associative and similarity-based processes in categorization decisions. *Memory & Cognition*, *25*, 625-640.
- Hampton, J. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition*, *65*, 137-165.
- Hampton, J. (2000). Concepts and prototypes. *Mind & Language*, *15*, 299-307.
- Hampton, J. (2006). Concepts as prototypes. En B. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 79-113). New York: Academic Press.
- Hull, C. (1920). Quantitative aspects of evolution of concepts: An experimental study. *Psychological Monographs*, *28*(1), 1-86.
- Katz, J., & Fodor, J. (1963). The structure of a semantic theory. *Language*, *39*, 170-210.
- Katz, J., & Postal, P. (1964). *An integrated theory of linguistic descriptions*. Cambridge, MA: MIT Press.
- Keil, F. (1983). Semantic inferences and the acquisition of word meaning. In T. Seiler, & W. Wannenmacher (Eds.), *Concept development and the development of word meaning* (pp. 103-124). Berlin: Springer-Verlag.
- Keil, F. (1986). The acquisition of natural kind and artifact terms. En W. Domopoulous, & A. Marras (Eds.), *Language learning and concept acquisition* (pp. 133-153). Norwood, NJ: Ablex.
- Keil, F. (1989a). *Concepts, Kinds and Cognitive Development*. Cambridge: MIT Press.
- Keil, F. (1989b). Spiders in the web of belief: the tangled relations between concepts and theories. *Mind & Language*, *4*, 43-50

- Keil, F. (1991). The emergence of theoretical beliefs as constraints on concepts. In S. Carey, & R. Gelman (Eds.), *The Epigenesis of Mind: Essays on Biology and Cognition* (pp. 237-256). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Keil, F. (2010). Hybrid Vigor and Conceptual Structure. *Behavioral and Brain Sciences*, 33, 215-216.
- Keil, F., & Wilson, R. (2000). The concept concept: The wayward path of cognitive science. *Mind & Language*, 15, 308-318.
- Keil, F., Smith, C., Simons, D., & Levin, D. (1998). Two dogmas of conceptual empiricism: implications for hybrid models of the structure of knowledge. *Cognition*, 65, 103-135.
- Keren, G., & Schul, Y. (2009). Two is not always better than one. *Perspectives on Psychological Science*, 4(6), 533-550.
- Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and Deliberate Judgments Are Based on Common Principles. *Psychological Review*, 118, 97-109.
- Kuhn, T. (1962). *The structure of scientific revolutions* (Vol. 58). Chicago: University of Chicago Press.
- Kunda, Z., Miller, D., & Claire, T. (1990). Combining Social Concepts: The Role of Causal Reasoning. *Cognitive Science*, 14, 551-578.
- Kushnir, T., Gopnik, A., Lucas, C., & Schulz, L. (2010). Inferring hidden causal structure. *Cognitive Science*, 34, 148-160.
- Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.
- Lalumera, E. (2010). Concepts are a functional kind. *Behavioral and Brain Sciences*, 33, 217-218.
- Landau, B. (2000). Concepts, the Lexicon and Acquisition: Fodor's New Challenge. *Mind & Language*, 15(2&3), 319-326.
- Laurence, S., & Margolis, E. (1999). Concepts and cognitive science. In E. Margolis, & S. Laurence (Eds.), *Concepts: Core readings* (pp. 3-81). Cambridge, MA: MIT Press.
- Lieberman, M. (2007). Social Cognitive Neuroscience: A Review of Core Processes. *Annual Review of Psychology*, 58, 259-289.
- Lucas, C., Bridgers, S., Griffiths, T., & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, 131, 284-299.
- Machery, E. (2005). Concepts are not a natural kind. *Philosophy of Science*, 72, 444-467.
- Machery, E. (2006). Two Dogmas of Neo-Empiricism. *Philosophy Compass*, 1(4), 398-412.

- Machery, E. (2007). 100 years of psychology of concepts: the theoretical notion of concept and its operationalization. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 38, 63-84.
- Machery, E. (2009). *Doing without Concepts*. Oxford: Oxford University Press.
- Machery, E. (2010). The heterogeneity of knowledge representation and the elimination of concept. *Behavioral and Brain Sciences*, 33, 231-239.
- Machery, E., & Seppälä, S. (2011). Against hybrid theories of concepts. *Anthropology & Philosophy*, 99-127.
- Mandler, J. (2004). *The Foundations of Mind: Origins of Conceptual Thought*. New York: Oxford University Press.
- Margolis, E., & Laurence, S. (2003). Concepts. En S. Stich, & T. Warfield (Eds.), *The Blackwell Guide to Philosophy of Mind* (pp. 190-213). Blackwell.
- Margolis, E., & Laurence, S. (2007). The Ontology of Concepts — Abstract Objects or Mental Representations? *Noûs*, 41(4), 561-593.
- Margolis, E., & Laurence, S. (2010). Concepts and theoretical unification. *Behavioral and Brain Sciences*, 33, 219-220.
- Margolis, E., & Laurence, S. (2014). *Concepts*. (E. N. Zalta, Ed.) The Stanford Encyclopedia of Philosophy (Spring 2014 Edition):
<http://plato.stanford.edu/archives/spr2014/entries/concepts/>
- Markman, A. (2011). Can Developmental Psychology Provide a Blueprint for the Study of Adult Cognition? *Behavioral and Brain Sciences*, 34(3), 140-141.
- Marques, J. F. (2006). Specialization and semantic organization: Evidence for multiple semantics linked to sensory modalities. *Memory & Cognition*, 34, 60-67.
- McCloskey, M., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition*, 6(4), 462-472.
- Medin, D., & Schaffer, M. (1978). Context Theory of Classification Learning. *Psychological Review*, 85, 207-238.
- Medin, D., & Shoben, E. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, 20, 158-190.
- Murphy, G. (2000). Explanatory concepts. In F. Keil, & R. Wilson (Eds.), *Explanation and cognition* (pp. 361-392). Cambridge, MA: MIT Press.
- Murphy, G. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

- Murphy, G., & Medin, D. (1999). The role of theories in conceptual coherence. En E. Margolis, & S. Laurence (Eds.), *Concepts: Core Readings* (pps. 425-458). Cambridge, MA: MIT Press.
- Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.
- Nosofsky, R. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology*, *115*(1), 39-57.
- Nosofsky, R. (1988). Exemplar-based accounts of relations between classification, recognition and typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*(4), 700-708.
- Nosofsky, R. (1992). Exemplars, prototypes and similarity rules. En A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning theory to connectionist theory: Essays in honor of William K. Estes* (Vol. 1, pp. 149-167). Hillsdale, NJ: Lawrence Erlbaum Associates
- Nosofsky, R., & Palmeri, T. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266-300.
- Nosofsky, R., Palmeri, T., & McKinley, S. (1994). Rule-Pus-Exception model of classification learning. *Psychological Review*, *101*, 53-79.
- Nosofsky, R., & Zaki, S. (2002). Exemplar and prototype models revisited: response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *28*(5), 924-940.
- Nosofsky, R., Cao, R., Cox, G., & Shiffrin, R. (2014). Familiarity and categorization processes in memory search. *Cognitive Psychology*, *75*, 97-129.
- Nosofsky, R., Cox, G., Cao, R., & Shiffrin, R. (2014). An exemplar-familiarity model predicts short-term and long-term probe recognition across diverse forms of memory search. *Journal of Experimental Psychology: Learning, Memory and Cognition*.
- Nosofsky, R., Little, D., Donkin, C., Little, D., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review*, *118*(2), 280-315.
- Osherson, D., & Smith, E. (1981). On the Adequacy of Prototype Theory as a Theory of Concepts. *Cognition*, *9*, 35-58.
- Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, *11*(6), 988-1010.

- Pecher, D. (2013). The perceptual representation of mental categories. In D. Reisberg (Ed.), *The Oxford Handbook of Cognitive Psychology* (pp. 358-373). Oxford: Oxford University Press.
- Pecher, D., Zeelenberg, R., & Barsalou, L. (2003). Verifying different-modality properties for concepts produces switching costs. *Psychological Science, 14*(2), 119-124.
- Pecher, D., Zeelenberg, R., & Barsalou, L. (2004). Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior activation. *Psychonomic Bulletin & Review, 11*(1), 164-167.
- Pinker, S., & Prince, A. (1999). The nature of human concepts: Evidence from an unusual source. En R. Jackendoff, P. Bloom, & K. Wynn (Eds.), *Language, logic and concepts* (pp. 221-261). Cambridge, MA: MIT Press.
- Prinz, J. (2002). *Furnishing the Mind: Concepts and their Perceptual Basis*. Cambridge, MA: MIT Press.
- Prinz, J. (2003). Author's response. *Metascience, 12*, 297-303.
- Prinz, J. (2004). Sensible ideas: a reply to Sarnecki and Markman and Stilwell. *Philosophical Psychology, 17*, 419-430.
- Prinz, J. (2005). The Return of Concept Empiricism. En H. Cohen, & C. Lefebvre (Eds.), *Categorization and Cognitive Science*. Elsevier.
- Putnam, H. (1975). The meaning of 'meaning'. En H. Putnam, *Philosophical Papers* (Vol. 2, pp. 215-271). Cambridge: Cambridge University Press.
- Reber, A. (1993). *Implicit Learning and Tacit Knowledge*. New York: Oxford University Press.
- Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science, 27*, 709-748.
- Rehder, B., & Burnett, R. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology, 50*, 264-314.
- Rehder, B., & Kim, S. W. (2006). How causal knowledge affects classification: a generative theory of categorization. *Journal of Experimental Psychology: Learning, Memory and Cognition, 32*, 659-683.
- Rehder, B., & Kim, S. W. (2010). Causal status and coherence in causal-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1171-1206.
- Rey, G. (2005). Philosophical analysis as cognitive psychology: the case of empty concepts. En H. Cohen, & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 71-89). Oxford: Elsevier.

- Rips, L. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Learning Behavior*, 14, 665-681
- Rips, L. (1989). Similarity, typicality, and categorization. En S. Vosniadou, & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21-59). Cambridge: Cambridge University Press.
- Rips, L., Shoben, E., & Smith, E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 12, 1-20.
- Robbins, P. (2002). How to Blunt the Sword of Compositionality. *Noûs*, 36(2), 313-334.
- Rosch, E. (1973). On the internal structure of perceptual and semantic categories. En T. Moore (Ed.), *Cognitive development and acquisition of language* (pp. 111-144). New York: Academic Press.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology*, 104, 192-233.
- Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.
- Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.
- Ryder, D. (2003). Making Sense of Empiricism? A Review of Jesse Prinz, *Furnishing the Mind*. *Metascience*, 12, 291-297.
- Sarnecki, J. (2004). The multimedia mind: an analysis on Prinz on concepts. *Philosophical Psychology*, 17(3), 403-418.
- Scarantino, A. (2010). Evidence of coordination as a cure for concept eliminativism. *Behavioral and Brain Sciences*, 33, 223-224.
- Schneider, S. (2009). The Nature of Symbols in the Language of Thought. *Mind and Language*, 24(4), 523-553.
- Schneider, S. (2011). *The language of thought: A new philosophical direction*. MIT Press
- Skidelsky, L. (en prensa). *Representaciones mentales: ¿Un eslabón entre el individuo y el mundo?* Buenos Aires: Catálogos.
- Slovan, S. (1996). The Empirical Case for Two Systems of Reasoning. *Psychological Bulletin*, 119, 3-22.
- Slovan, S. (2002). Two systems of reasoning. En T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 379-398). Cambridge: University Press.

- Sloman, S. (2014). Two Systems of Reasoning: An update. In J. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 69-79). New York: Guilford Press.
- Smith, D., & Minda, J. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology*, *24*(6), 1411-1436.
- Smith, E., & Medin, D. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.
- Smith, E., & Osherson, D. (1989). Similarity and decision making. In S. Vosniadou, & A. Ortony (Eds.), *Similarity and Analogical Reasoning*. (pp. 60-75). New York: Cambridge University Press
- Smith, E., Osherson, D., Rips, L., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science*, *12*, 485-527.
- Smith, J. D. (2002). Exemplar theory's predicted typicality gradient can be tested and disconfirmed. *Psychological Science*, *13*, 437-442
- Smoke, K. (1932). An objective study of concept formation. *Psychological Monographs*, *42*, 1-46.
- Spelke, E., & Kinzler, K. (2007). Core knowledge. *Developmental Science*, *10*(1), 89-96.
- Spelke, E., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*, 605-632.
- Stanfield, R., & Zwaan, R. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153-156.
- Stanovich, K. (2004). *The Robot's Rebellion*. Chicago: The University of Chicago Press.
- Stanovich, K., & Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind & Society*, *11*, 3-13.
- Sun, R. (2002). *Duality of the mind*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Sun, R., Slusarz, P., & Terry, C. (2005). The Interaction of the Explicit and the Implicit in Skill Learning: A Dual-Process Approach. *Psychological Review*, *112*, 159-192.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327-352
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293-315.
- van Dantzig, S., Pecher, D., Zeelenberg, R., & Barsalou, L. (2008). Perceptual Processing Affects Conceptual Processing. *Cognitive Science*, (*32*), 579-590.
- Vermeulen, N., Corneille, O., & Niedenthal, P. (2008). Sensory load incurs conceptual processing costs. *Cognition*(109), 287-294.

- Weinberg, J. (2003). Making Sense of Empiricism? A Review of Jesse Prinz, Furnishing the Mind. *Metascience*, 12, 279–285.
- Weiskopf, D. (2009a). The plurality of concepts. *Synthese*, 169, 145-173.
- Weiskopf, D. (2009b). Atomism, pluralism, and conceptual content. *Philosophical and Phenomenological Research*, 79, 130-162.
- Weiskopf, D. (2009c). The theory theory of concepts. *Internet Encyclopedia of Philosophy*.
<http://www.iep.utm.edu/th-th-co/>
- Weiskopf, D. (2010). The theoretical indispensability of concepts. *Behavioral and Brain Sciences*, 33, 228-229.
- Whorf, B. (1956). *Language, thought and reality*. Cambridge, Ma: MIT Press
- Wittgenstein, L. (2002). *Investigaciones filosóficas*. México: UNAM - Crítica.
- Xu, F., & Spelke, E. (2000). Large number discrimination in 6-month old infants. *Cognition*, 74, B1-B11.