



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE FILOSOFÍA Y LETRAS

El debate en torno a las explicaciones
reductivas y no reductivas de la conciencia
fenoménica

T E S I S

Que para obtener el título de

Licenciado en Filosofía

P R E S E N T A

Jovany Axel Tovar Carmona

DIRECTOR DE TESIS

Dr. Miguel Ángel Sebastián González

Cd. Universitaria. Cd. Mx. 2019



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

A la memoria de Aldo Carmona y Flavio Carmona.

Por los ejemplos de dedicación, humildad y compromiso

Y porque los extraño mucho y nada me gustaría más que seguir compartiendo con ellos

Agradecimientos

Si bien con esta tesis termino oficialmente la licenciatura, para mí el trabajo realizado es en realidad el comienzo de mi formación filosófica. No pude tener un mejor asesor en ambos aspectos que el Dr. Miguel Ángel Sebastián, quien no sólo ha dirigido este trabajo con rigor y paciencia, sino que sigue enseñándome temas de filosofía de la mente, de metodología y argumentación filosófica; así como inspirándome a continuar en la filosofía. Por todo ello, infinitas gracias.

Además, merecen un gran reconocimiento los lectores de esta tesis, por su tiempo y valiosos comentarios: Dr. Eduardo García Ramírez; Dr. Cristian Gutiérrez, de quien además aprendí mucho sobre cómo hacer un trabajo de titulación durante su seminario de tesis; Dr. Axel Barceló, cuyas clases, por cierto, fueron algunas de mis favoritas durante la licenciatura; y al Mtro. Luis Graillet, quien además me permitió desarrollar algunas habilidades docentes durante mi servicio social. ¡Gracias!

Por otro lado, quiero agradecer al Instituto de investigaciones filosóficas, como entidad abstracta y como comunidad, por permitirme formar parte del programa de estudiantes asociados. Durante este programa he podido discutir algunos de los temas que más me interesan con varios de sus miembros, con quienes además he forjado una amistad muy grata, especialmente con los miembros del Aula 3, recordándome constantemente que la filosofía es una actividad colectiva.

También porque proveer de un espacio de trabajo y de acceso a una biblioteca especializada en filosofía analítica en una universidad tan sobrepoblada es realmente invaluable.

En otra dimensión, quizá más importante que la académica, tengo que reconocer y agradecer a diversas personas sin cuyo apoyo no habría podido terminar nunca una carrera universitaria. En primer lugar, mi mamá, Aidé Carmona, quien no sólo ha sido terriblemente paciente y amorosa, sino que fue mecenas de todos mis estudios; Mamá, hiciste algo increíble y espero remunerarlo pronto. En segundo lugar, mi hermana, porque sin las risas espontáneas compartidas y sus abrazos, no habría podido lidiar con las difíciles situaciones que hemos tenido que afrontar. Las amo a ambas <3.

Además, debo agradecer a mis titos, quienes estuvieron pendientes de mi existencia y de mi formación día con día y compartieron con mucha alegría varios de mis pequeños logros, recordándome que todo esto vale la pena.

Agradezco también a mis amigos. A Ana, Andrés y Christian, quienes además de preocuparse por mi formación, han acompañado mis crisis y mis quejas, pero también me han dado muchos de los momentos más bellos y favoritos de mi vida. A Sofía, porque, además de ser partícipe de lo anterior, su compañía durante las tardes de café y panqués veganos en la biblioteca de la facultad, hizo que el inicio de este trabajo fuera una tarea más sencilla y llevadera.

Sin todos los mencionados nada de esto hubiera sido posible. Muy pocas veces es necesario decir lo obvio, pero para concluir esta sección, es ineludible señalar que todos los errores que pudiese haber en este trabajo son completamente causa mía y no de ninguno de los mencionados (que tampoco hacen milagros).

Índice

Introducción	6
Capítulo 1. Conciencia fenoménica	12
1.1 Conciencia y conciencia fenoménica	12
1.2 La cuestión ontológica	18
1.2.1 Fisicalismo	19
1.2.2 Dualismo	22
1.3 La cuestión explicativa	23
Capítulo 2. Explicaciones Reductivas y análisis conceptual	26
2.1 Explicaciones Reductivas	26
2.2 La tesis de la implicación <i>a priori</i>	32
2.3 Análisis conceptual y significado	35
2.3.1 Bidimensionalismo	38
2.4 Explicaciones reductivas y análisis conceptual	40
Capítulo 3. Las explicaciones reductivas de la conciencia fenoménica	45
3.1 El argumento de la concebibilidad: Zombies	46
3.2 El argumento del conocimiento: Murciélagos y neurocientíficas presas	50
3.3 Conceptos fenoménicos y psicológicos: Dos formas de referir a la mente	54
Capítulo 4: Las explicaciones no reductivas de la conciencia fenoménica	59
4.1. Block y Stalnaker en contra del análisis conceptual	60
4.2 El modelo nageliano de explicación reductiva	63
4.3 Identidades a posteriori y explicaciones reductivas (Block y Stalnaker otra vez)	67
4.4 Las explicaciones no reductivas de la conciencia fenoménica	71
Conclusiones	73
Referencias	77

Introducción

No es controvertido pensar que existe una relación íntima entre nuestros estados mentales y nuestros estados físicos, específicamente estados cerebrales y de nuestro sistema nervioso. Esta idea es apoyada y reconocida por algunos enfoques de la psicología y otras ciencias cognitivas, que cada vez se acercan más a las neurociencias y a los estudios empíricos del cerebro en busca de aumentar el conocimiento acerca de la mente. Pero también nuestras intuiciones y prácticas cotidianas parecen reconocer y operar a partir de esta idea: si sentimos un dolor, por dar un ejemplo, tomamos un analgésico que actúa a nivel de las estructuras químicas cerebrales para detener la sensación.

En el contexto de esta idea surgen los programas reduccionistas de la mente, de entre los cuales, a grandes rasgos, se pueden identificar dos enfoques principales. El primero de ellos es un enfoque ontológico: los estados mentales se proponen reductibles a lo físico cuando se considera que no existe nada más por encima de los estados físicos; a grandes rasgos, nuestros estados mentales no serían otra cosa que estados cerebrales-corporales. El segundo enfoque es una propuesta de carácter explicativo, que propone que los estados mentales son explicables completamente a partir de términos físicos.

Por supuesto, cualquiera de estos proyectos podría resultar falso. Pero tal como el debate en torno al problema mente-cuerpo se desarrolla actualmente, considero que su falsedad sería más bien una conclusión que el punto de partida.

Estos dos enfoques suelen ir de la mano. Particularmente el segundo se inscribe dentro del primero, pues si la mente resulta ser algo más que estados físicos, explicarla a partir de éstos no tiene mucho sentido. Sin embargo, mi intención aquí no es profundizar en su relación exacta. En esta tesis me concentraré en el segundo proyecto, aquel acerca de ofrecer explicaciones de carácter reductivo para los estados mentales.

En general, resulta importante evaluar la posibilidad de este tipo de explicaciones pues éstas se pueden ser parte importante para el proceder científico, ya que mediante ellas se puede hacer inteligible la ocurrencia de un fenómeno a la luz de otros que conocemos mejor y que creemos tienen una relación de dependencia o identidad con los primeros. Además, proveer explicaciones de algunos sucesos pertenecientes a un dominio científico particular en los términos de una ciencia diferente, comúnmente de mayor legitimidad y aceptación, se ha considerado parte de un proyecto de unificación científica que otorga legitimidad a las ciencias particulares y que al mismo tiempo ha pasado a ser guía metodológica para muchas de ellas.

Dado que las ciencias cognitivas son un entramado de disciplinas de relativamente reciente aparición y cuyo desarrollo recién comienza, es particularmente pertinente evaluar la posibilidad de explicar algunos fenómenos y sucesos mentales en términos de los estados físicos que consideramos subyacentes, pues a estos últimos tenemos un mejor acceso epistémico desde una perspectiva de tercera persona.

Por otro lado, existen diferencias importantes entre nuestros estados mentales. Por ejemplo, una creencia política es distinta a la sensación de un calambre, pero a ambos los consideramos, tanto intuitivamente como en debates filosóficos y psicológicos, estados mentales.

Así, es necesario aclarar que en esta tesis me enfoco solamente en los denominados estados fenoménicamente conscientes. En palabras de Nagel (1974), un estado mental es fenoménicamente consciente cuando hay algo *que es como estar* en ese estado para un organismo. A grandes rasgos, podemos entender los estados fenoménicamente conscientes como aquellos que se sienten de algún modo.

Explicar la presencia de la conciencia fenoménica en términos físicos es, entonces, dar cuenta de porqué cierto carácter fenoménico y no otro o ninguno, acompaña a un estado físico-cerebral y no a otro, usando fundamentalmente términos físicos. Esto es lo que David Chalmers (1996) ha denominado el problema duro de la conciencia (*the hard problem of consciousness*) y es uno de los temas más discutidos en filosofía de la mente y de las ciencias cognitivas actualmente. De no poder dar respuesta satisfactoria, estaremos frente a lo que Levine (1983) denominó una brecha explicativa.

En consiguiente, esta tesis está guiada por la pregunta acerca del tipo de explicaciones que podemos dar de la presencia de la conciencia fenoménica, es decir, si estas pueden ser reductivas o no. Mi objetivo particular es mostrar que tenemos buenas razones para pensar que esto no es el caso, que efectivamente existe un problema duro de la conciencia, es decir, que no podemos dar cuenta de la ocurrencia de los estados fenoménicos usando términos puramente físicos.

Es cierto que esta conclusión es poco novedosa (aunque por otro lado, tampoco ha logrado una completa aceptación). Pero, considero que el valor de esta tesis no está meramente en la conclusión que suscribo, sino en dos puntos principales: el primero de ellos es que aunque me inclino hacia un lado de debate, un objetivo secundario es presentar con detalle y claridad los argumentos y supuestos de éste, por lo que esta tesis podría tener un gran valor monográfico – como he tratado de reflejar desde su título- para quienes recién se inician en debate acerca de la conciencia.

En segundo lugar, algo que comúnmente está ausente en el debate en torno a las explicaciones reductivas de la conciencia fenoménica es especificar qué se entiende por explicación y qué caracteriza a una explicación que se considera reductiva. Por tal motivo, en este trabajo he decidido limitar la discusión a un modelo particular de explicación científica: el modelo nomológico deductivo. Considero que de hecho este es el modelo implícito en gran parte del debate, pero aún si esto no fuere el caso, muestro como se aplicarían los términos generales de la discusión a este modelo particular. Con lo cual, intento aportar mayor claridad al debate.

Así, la metodología general de esta tesis será defender y relacionar un requisito necesario para una explicación reductiva dentro de este modelo particular con un requisito general para la explicación reductiva esbozado en el debate en torno a la conciencia. Mostraré después los argumentos en contra de la satisfacción de estos requisitos para el caso de la conciencia, concluyendo que ésta no puede ser sujeto de explicaciones reductivas. Adicionalmente, expondré dos propuestas de explicación reductiva alternativa, bajo las cuales la conciencia podría explicarse, pero suscribiré una crítica que las señala como no reductivas, en particular porque no cumplen con el requisito solicitado y al no hacerlo enfrentan desventajas epistémicas. Este último paso aportará también, aunque indirectamente y de forma quizás no concluyente, a la defensa de los requisitos solicitados.

Así, esta tesis se divide en cuatro capítulos cortos. El primero de ellos está dedicado principalmente a exponer la noción de conciencia fenoménica, distinguiéndola de otras acepciones del término “conciencia” usadas en psicología, en filosofía o en el lenguaje cotidiano, con el fin de evitar confusiones. Además, presento a grandes rasgos el marco ontológico, el fisicalismo, dentro del cual se inscribe la discusión que trato, y con ello algunos detalles acerca de lo que las propiedades y términos físicos son. Además, adelanto brevemente porque esta noción de conciencia es particularmente problemática.

El segundo capítulo lo dedico a exponer la segunda noción importante en el título de este trabajo: las explicaciones reductivas. Comienzo este capítulo exponiendo el modelo nomológico deductivo de explicación científica, en el que centraré la discusión. Posteriormente expondré un requisito, (R), que debe cumplir una explicación dentro de este modelo para considerarse como reductiva y un problema de tensión aparente entre el requisito y el modelo. Finalmente, relacionaré este requisito con la tesis de la implicación *a priori*, que se postula como requisito necesario para las explicaciones reductivas (comúnmente ya en el debate acerca de la conciencia) y expondré como la solución al problema de tensión, y lo que fundamentalmente permite una explicación reductiva, son consideraciones semánticas de los términos involucrados, esto es, un tipo de análisis conceptual.

En el tercer capítulo expondré los argumentos más famosos en contra de las explicaciones reductivas de la conciencia: el argumento de la concebibilidad y el argumento del conocimiento, que se dirigen directamente contra la tesis de la implicación *a priori*. Mencionaré además que subyacente a estos argumentos se encuentra una distinción entre los conceptos fenoménicos y psicológicos, los primeros de los cuales no son susceptibles del análisis conceptual requerido. Ejemplificaré estos argumentos en el modelo nomológico deductivo, mostrando como al no satisfacer la tesis de la implicación *a priori* tampoco se satisface el requisito (R). Con esto, inclinaré la balanza en contra de la posibilidad de ofrecer explicaciones reductivas de la conciencia en el modelo ofrecido.

Finalmente, en el cuarto capítulo expondré una famosa crítica a la necesidad del análisis conceptual para las explicaciones reductivas, evaluando también dos famosas alternativas, aun dentro del modelo nomológico deductivo, dentro de las cuales podría darse cuenta de la conciencia fenoménica: las leyes puente y las identidades *a posteriori*. Suscribiré una respuesta a estas alternativas que las señala como no reductivas, y que, al no serlo, carecen de las ventajas

epistémicas de las explicaciones que sí lo son. Así, mencionaré que aunque puede “explicarse” la conciencia en estos modelos, estas explicaciones no son reductivas, pues además de los términos físicos, requieren incorporar los términos fenoménicos como básicos en el sistema explicativo.

Vale la pena señalar, como nota metodológica, que la filosofía de la mente es una disciplina amplia, que se relaciona fuertemente con otras áreas de la filosofía (especialmente con la filosofía de la ciencia, del lenguaje y con la metafísica), y muchas otras veces la distinción entre ellas ni siquiera es tan tajante.

La discusión que presento no es la excepción. Algunos de los términos sobre los que se construye no son incontrovertidos, como es de esperarse en filosofía, y en muchos casos pertenecen a distintas áreas. Por razones metodológicas no puedo contestar todas las posibles objeciones o detallar cada controversia, o esta tesis tendría un ciento de páginas más y hubiese tardado, al menos, un par de años extra para lograrlo. No obstante, en las notas al pie de página he intentado referir a los autores principales en que las discusiones se desarrollan, hasta donde mi neófita y restringida experiencia filosófica me ha permitido reconocer, y trato de señalar los límites de mi investigación constantemente dentro del cuerpo del texto.

Capítulo 1:

Conciencia fenoménica.

Como ya se ha dicho en la introducción, el objetivo de esta tesis es evaluar las explicaciones reductivas de los estados fenoménicamente conscientes. Dedicaré este capítulo a exponer la noción de conciencia fenoménica y de estados fenoménicamente conscientes, fundamentalmente en la sección 1.1. Mi objetivo es distinguir esta noción de otras nociones de conciencia usuales en la literatura y en la vida cotidiana para evitar confusiones.

Además, expondré en la sección 1.2 el marco ontológico en los que se inscribe el debate en torno a las explicaciones reductivas de la conciencia fenoménica, el fisicalismo. Esto servirá también para exponer qué es una propiedad física. Además, para cerrar el capítulo, comenzaré a bosquejar la cuestión explicativa, en la que se centra esta tesis, en la sección 1.3.

1.1 Conciencia y conciencia fenoménica

De forma intuitiva nuestros estados mentales difieren radicalmente unos de otros, particularmente por la forma en que los experimentamos, o bien porque son acerca de distintas cosas. No hay mucho que discutir al respecto. Mi creencia de que me encuentro en el aula 3 del Instituto de investigaciones filosóficas, mi deseo de salir a buscar una taza café o la sensación de

dolor del calambre que me despertó hace poco a media noche, por poner algunos ejemplos, son todos estados mentales que difieren entre sí. Además, a algunos de nuestros estados mentales atribuimos la propiedad de ser conscientes¹; es precisamente la conciencia el tema general de este trabajo.

Una de las citas más recurrentes en filosofía de la mente, al menos como epígrafe, se la debemos a T. H. Huxley (1866): “Cómo algo tan extraordinario como un estado consciente surge como resultado de algo como la irritación del tejido nervioso, es tan inexplicable como la aparición del genio cuando Aladín frotó su lámpara” (Citado en Block (1995; 231) traducción mía). Permitiéndose ciertas licencias literarias, Huxley expresa lo enigmática y central que consideramos es la conciencia en el problema mente-cuerpo.

Aunque intuitiva o preteóricamente coincidamos con él, no siempre es muy claro a qué nos referimos por “conciencia” o por “estado mental consciente”. Mi objetivo en esta sección es aclarar qué es la *conciencia fenoménica* y qué estados son *fenoménicamente conscientes*, el tipo de estados mentales en los que esta tesis se centra, y distinguirlos tanto de otros estados mentales como de otras nociones de “conciencia”.

No son pocos los casos en los que algunas palabras de uso cotidiano se usan en debates técnicos con un sentido poco relacionado al que les damos en el día a día, o, inversamente, en la cotidianidad se usan términos técnicos de forma poco rigurosa. En los contextos no filosóficos o psicológicos, “conciencia” se usa con tal laxitud que bajo esta etiqueta caen distintos fenómenos que podrían no tener mucha relevancia filosófica, por ejemplo, la conciencia entendida como una

¹ Suscribo la tesis general de que no todos nuestros estados mentales son conscientes bajo cualquier interpretación no trivial de “conciencia”. En la posición contraria, una perspectiva que podemos llamar “cartesiana”, en la que todo estado mental es consciente, la noción de conciencia implicada es poco clara además no facilita distinciones relevantes. (Rosenthal (1996) discute detalladamente estas posturas). Más adelante se verá que al menos para la noción de conciencia que me interesa, hay efectivamente estados mentales no conscientes.

especie de guía o responsabilidad de nuestros actos morales –algunas veces en forma de grillo- o políticos. Es importante, entonces, distinguir cualquier acepción técnica de “conciencia” de los usos cotidianos del término.

Buena parte de la filosofía de la mente se ha encargado de establecer distinciones conceptuales entre tipos de estados mentales, algunas veces incluso con miras a mejorar la práctica científica empírica². Aunque quizá no existe una taxonomía definitiva, pues gran parte de las clasificaciones dependen de los objetivos particulares de los autores que las proponen, y muchas veces se puede ser aún más fino, hay ciertas nociones compartidas en la literatura filosófica. En lo siguiente, pretendo presentar una clasificación no controvertida que sea sólo lo suficiente fina para mis propósitos, esto es, que evite confundir la noción de conciencia fenoménica con otros tipos de conciencia y que permita ver qué tipo de estados mentales son fenoménicamente conscientes. Las consideraciones siguientes las retomo particularmente de Block (1995), Rosenthal (1986), Lycan (1996,1-13) y Carruthers (2000, 9-22) (De entre estos, son los dos últimos quienes realizan clasificaciones más detalladas), aunque están presentes en gran parte de la literatura.

A veces, al decir que alguien es consciente queremos decir que se encuentra en un estado de atención o alerta, con una vida mental compleja capaz de procesar información y responder a los estímulos ambientales. Esto es, aproximadamente, lo opuesto a estar dormido, anestesiado o bajo el efecto de alguna sustancia depresora. Este es el sentido de conciencia que usamos en frases del tipo “mi hermana estuvo consciente durante su cirugía, pues sólo usaron anestesia local”. En este caso decimos que es *el organismo* el que se encuentra consciente, es decir, la conciencia es una propiedad que predicamos de la totalidad de un sistema (Lycan, 1996; 2). Además esta noción

² Véase, por ejemplo, Block (1995) en donde se propone que una confusión entre dos tipos de conciencia (conciencia de acceso y conciencia fenoménica) ha llevado a conclusiones falaces en algunas investigaciones empíricas (el ejemplo ofrecido es la investigación en torno a la presencia de conciencia en fenómenos de *blindsight*).

parece ser gradual (se puede estar más o menos consciente) y algunas veces la usamos en un sentido más disposicional que actual. Esta es una primera acepción de conciencia, aunque todavía muy general y de uso más cotidiano que técnico.

Otras veces somos un poco más exigentes y pedimos que la conciencia de un organismo esté necesariamente *dirigida a algo*. Por ejemplo, ahora me encuentro en un estado de vigilia, pero mi conciencia no recae simplemente en ello, sino en mi percatación de algo: soy consciente *de que* ahora me encuentro despierto, *de que* finalmente estoy bebiendo café, y soy también consciente *de que* no puedo resolver el problema mente-cuerpo, etc. Esta es una noción de conciencia *transitiva*, pues está dirigida a algo (Lycan, 1996; 3). Al igual que la noción anterior es más preteórica que técnica y a veces no suele distinguirse de nuestro uso, también preteórico, del término “conocimiento”.

La transitividad, por otro lado, a veces es predicada también de estados mentales, no sólo de organismos. De tal forma, algunos de nuestros estados mentales están dirigidos a algo, o tienen algún contenido. Por ejemplo, mi conocimiento acerca del lugar en que me encuentro tiene como contenido el aula 3 del IIFs, mi deseo de café está dirigido al café, etc. Estos estados mentales se conocen como estados intencionales. Estados intencionales típicos son las creencias, los deseos, el saber, el imaginar, etc. pues no podemos simplemente creer, desear, saber o imaginar, sino que necesariamente estos estados son acerca de algo.

Aunque en los debates especializados no suele considerarse que los estados intencionales por sí mismos sean conscientes, hay otros estados mentales a los que sí atribuimos conciencia. En primer lugar, solemos decir que un estado mental es consciente cuando nos percatamos de estar en él, es decir, cuando tenemos algún pensamiento o percepción de orden superior que lo tiene como contenido. Este estado de segundo orden es un estado intencional cuyo contenido es otro estado mental, pero es el de primer orden al que atribuimos la propiedad de ser consciente.

Podemos referirnos a este tipo de conciencia como algo aproximado a la “percatación consciente” (*awareness*) (Rosenthal, 1986). Esta, sin embargo, no es la noción de estado mental consciente que me interesa tratar aquí.

Finalmente, hay un tipo de estado mental a los que podemos atribuir un sentido de conciencia que, se ha dicho (por ejemplo por Block, 1995), no es necesariamente transitivo: la conciencia fenoménica³. Como ya se anticipó es esta la noción que me interesa en este trabajo. A grandes rasgos los estados fenoménicamente conscientes son estados que se sienten de algún modo; en palabras de Thomas Nagel, “un organismo tiene estados [fenoménicamente] conscientes si y sólo si hay algo que es como es ser [*Something that it is like to be*] ese organismo [al estar en ese estado]” (1974, 436; traducción mía).

Esta es, quizá, la noción de conciencia más enigmática. Resulta tan problemática que cuesta trabajo ofrecer una definición precisa de la misma, incluso la definición técnica de Nagel no parece inmediatamente clara. La mayoría de los autores (Block (1995), Chalmers (1996), por poner sólo un par de ejemplos) al caracterizar la noción de conciencia fenoménica, dada la dificultad de definirla, señalan que la mejor forma de entenderla es a través de ejemplos y sinónimos. A falta de una mejor opción, seguiré ese camino.

Podemos pensar a los estados fenoménicamente conscientes como estados mentales que tienen alguna sensación, es decir, se sienten de alguna forma. Estados paradigmáticamente fenoménicos son las experiencias, específicamente las sensaciones obtenidas de ellas. Las sensaciones de los colores blanco y azul de la pantalla del computador en que escribo esto, el sabor y el olor del café que finalmente conseguí, los sonidos de la música que escucho en segundo plano, etc. Pero también son estados fenoménicos otras sensaciones corporales, como los dolores, los

³ Otro uso común de “conciencia” puede referirse a algún tipo de autoconciencia, pero esta es una noción mucho más complicada de exponer ya que tampoco es por sí mismo un término de uso unívoco. Además de la noción de percatación, autoconciencia requiere que ésta esté dirigida de algún modo al propio sujeto.

orgasmos, los calambres, las cosquillas o los adormecimientos de extremidades después de un tiempo sin moverlas. En todos estos estados *hay algo que es como* estar en ellos. Por esta razón, en la literatura se suelen usar términos como “experiencias conscientes” o “experiencias fenoménicas” o simplemente “experiencias” como sinónimos aproximados de “estados fenoménicamente conscientes”.

Es cierto que muchos otros estados mentales tienen algún componente fenoménico. Las emociones, como el miedo o la tristeza, por ejemplo, aunque no se agotan en su fenomenología⁴, poseen un componente sensitivo; los sentimientos pueden tener un aspecto fenoménico también, y quizás también otros estados como los partícipes en la imaginación, la memoria, el aprendizaje, u otros más complejos, lo tengan. Considero que con lo dicho y con los ejemplos dados, queda claro de qué se habla al hablar de estados fenoménicamente conscientes.

El consenso en filosofía de la mente es que podemos caracterizar cualquier estado mental mediante la intencionalidad o la conciencia fenoménica, sin excluir una mezcla de ambas (por ejemplo, Rosenthal (1986), Block (1995), Chalmers (1996)). De esta forma algunos estados, como los preceptuales, son una mezcla de intencionalidad y fenomenología, mientras que otros, como muchas creencias, muy probablemente sólo exhiben intencionalidad, es decir, son estados mentales no conscientes (en este sentido particular de conciencia).

No es mi intención profundizar en la relación entre estados intencionales y fenoménicos, pues esta es una cuestión no exenta de polémica⁵, pero en principio podría preguntarse si estados

⁴ Cuando digo que un estado mental exhibe alguna fenomenología, me refiero solo a que es un estado fenoménicamente consciente. No debe tomarse el término “fenomenología” de forma más fuerte que eso, en particular no debe pensarse que está relacionado con la escuela y método fenomenológicos.

⁵ Algunos, como Brentano (Véase Huemer, 2018) consideran que la intencionalidad es la marca distintiva de cualquier estado mental, incluidos por tanto los fenoménicos. Recientemente el representacionalismo pretende dar cuenta de los estados fenoménicos en términos representacionales-intencionales (Tye, 1995; 2009), recuperando en un sentido la tesis de Brentano. Algunos autores como Block (1995; 1990) apuntan a que los estados fenoménicos no son en modo alguno intencionales. Pero, más allá de si todo estado fenoménico es intencional, parece haber problemas para caracterizar y explicar la fenomenología

en los que parecen confluír la intencionalidad y la fenomenología son un solo estado con propiedades distintas (fenoménicas, el cómo se siente estar en ellos; e intencionales, acerca del contenido que tienen y cómo lo representan) o dos estados distintos, uno fenoménico y otro intencional, que coocurren (incluso sistemáticamente).

Existe, sin embargo, en ambos casos, una clara diferencia conceptual entre la caracterización fenoménica e intencional de los estados mentales. Por distinción conceptual me refiero, a grandes rasgos, a que es posible imaginar consistentemente, situaciones en las que un estado fenoménico pueda presentarse sin el componente intencional que tradicionalmente lo acompaña, o viceversa. Es entonces la presencia del carácter fenoménico, o de propiedades fenoménicas, es decir, que haya algo como estar en ese estado, lo que hace a un estado mental fenoménicamente consciente (con independencia de si este es intencional también o no) y es la presencia de este carácter la que me interesa en este trabajo.

1.2 La cuestión ontológica

Se ha mencionado tangencialmente que los estados fenoménicamente conscientes resultan particularmente problemáticos, pero no se ha dicho nada acerca de porqué esto es así. Los estados fenoménicos presentan dificultades para las posturas fisicalistas, aquellas que sostienen que todo lo que existe es físico, pues es difícil dar cuenta de su ocurrencia en términos físicos, como expondré en esta tesis. Además, muchos de los argumentos en contra del fisicalismo parten de esta conclusión explicativa para intentar concluir una brecha ontológica (Chalmers, 1996, 2002a; 2009) (es decir,

completamente en esos términos (Chalmers, 2004). Por otro lado, no sé de alguien que sostenga que la fenomenología es fundamento de la intencionalidad (en sentido ontológico o explicativo).

que los estados fenoménicos tienen una existencia no física), aunque el paso no es inmediato ((Loar, (1990); Stoljar, (2005), Hill (2002) entre muchos otros, resisten esta conclusión).

Aunque la discusión en la que me concentraré en este trabajo es acerca las explicaciones, tomando distancia de sus consecuencias ontológicas, vale la pena explicar brevemente en qué consiste el fisicalismo, pues es éste el marco ontológico en el que el debate por las explicaciones reductivas se inscribe. Además, presentar el fisicalismo permitirá caracterizar qué es una propiedad física y, con ello, qué son los términos físicos, lo que será relevante posteriormente. Esto es lo que haré en la siguiente sección.

1.2.1 Fisicalismo

El fisicalismo es la tesis metafísica que sostiene que todo lo que existe en el mundo es físico. Lo que “físico” significa es, como todo en filosofía, un asunto debatible, pero existen dos interpretaciones principales. La primera de ellas, por la que aquí me inclinaré, concibe que algo es una propiedad física si es una propiedad de la que la física da cuenta. La segunda interpretación, define las propiedades físicas como aquellas propiedades intrínsecas de los objetos físicos paradigmáticos. Estas dos versiones no seleccionan las mismas propiedades, pues la física da cuenta de propiedades relacionales, no intrínsecas. (Stoljar, 2001 explica la distinción con detalle, aunque él se inclina por la segunda caracterización).⁶

Ambas caracterizaciones tienen inconvenientes, algunos de los cuales Stoljar (2001) mismo aclara. Aquí me inclino por la primera caracterización, porque considero existe un consenso en discutir sobre esta interpretación y es menos problemática a nivel de acceso epistémico. Si las ciencias dan cuenta de propiedades relacionales, ¿bajo qué medios epistémicos podríamos saber

⁶ Considerar las propiedades físicas como propiedades intrínsecas de los objetos físicos paradigmáticos colapsa las posturas fisicalistas con posturas (proto) panpsiquistas. Véase la nota 10.

cuáles son las propiedades intrínsecas de los objetos?, además, ¿cuáles serían los objetos físicos paradigmáticos? Así, la segunda interpretación, en mi opinión, es aún más especulativa.

El problema más grande al que se enfrenta la interpretación de propiedad física a partir de las ciencias físicas es el famoso dilema de Hempel. De acuerdo con un cuerno del dilema, si lo que se considera como ciencia física es la física actual, el fisicalismo sería falso, pues es altamente probable que la física actual resulte falsa a la luz de una física futura. Por otro lado, si lo que se considera como la ciencia física relevante es precisamente una ciencia física futura última, la formulación del fisicalismo podría ser trivial o ininteligible.

Comúnmente se acepta que existe una salida al dilema de Hempel. Respecto al primer cuerno del dilema, se argumenta que la ciencia física del futuro complementaría, más que falsear, la ciencia actual, lo que no haría al fisicalismo falso. Respecto al segundo cuerno, se ha argumentado que es poco probable que la ciencia física del futuro incluya como básicas propiedades de un tipo tan diametralmente diferentes a las formuladas por la ciencia actual, con lo que el fisicalismo no sería trivial. Además, podemos hacernos una idea del tipo de propiedades que postularía una física futura a partir del tipo de propiedades de la física actual (Cf. Jackson, 1998; 6-7). Como mi objetivo no es discutir con detalle estos problemas, suscribiré también estas salidas como opciones viables.

Sin embargo, siguiendo la definición dada, no todas las propiedades y hechos del mundo serían físicos. Por ejemplo, de hechos mentales, económicos o sociales, no da cuenta directamente ninguna teoría física. Así, para no eliminar estos sucesos de nuestra ontología, podemos refinar nuestra definición, considerando que una propiedad es física si y sólo si es una propiedad de la que da cuenta alguna teoría física o mantiene algún tipo de relación adecuada con alguna de ellas. (Stoljar, 2001)

Ahora bien, se ha propuesto a la superveniencia/sobreveniencia global metafísica como la relación adecuada entre propiedades para caracterizar al fisicalismo. Esta tesis caracteriza un tipo

de dependencia de unas propiedades sobre otras a través de las relaciones de invariancia y covarianza entre ellas. De acuerdo con esta, una propiedad B superviene en una propiedad A si y solo si dos mundos metafísicamente posibles que son A -idénticos (idénticos respecto a la instanciación de propiedades A), serán también B -idénticos.⁷ Una forma alternativa de representar esta relación es a través del condicional $\Box (P \rightarrow M)$, en donde el operador modal es irrestricto a mundos metafísicamente posibles accesibles desde el mundo actual, P es una descripción completa del mundo en términos de las teorías físicas y M es una verdad macroscópica cualquiera (Jackson (1993, 1998; 1-28) llama a este condicional la tesis de la entrada por implicación (*the entry by entailment*), en dichas obras se encuentra expuesta también su equivalencia con la tesis de la superveniencia).

Sostener el fisicalismo es comprometerse con que toda propiedad es física o sobreviene metafísicamente en algunas de ellas. No es inmediatamente claro que sucesos sociales, políticos, morales, matemáticos, entre otros, resulten ser físicos, como sostendría el fisicalista; pero para los propósitos de esta tesis no es necesario comprometerse con algo tan fuerte, basta sostener que al menos respecto a lo mental nos inclinamos a un tipo de “fisicalismo”, es decir que al menos las propiedades mentales supervienen metafísicamente en las físicas.

Esta idea resulta más aceptable. Véase, como ejemplo, que la psicología cada vez toma más en cuenta los datos cerebrales, e investiga qué pueden decir estos sobre nuestros estados y procesos mentales, logrando resultados significativos. Pero también en la cotidianidad actuamos

⁷ Existen distintas versiones de la relación de superveniencia que no suelen considerarse suficientemente fuertes para caracterizar al fisicalismo. Por ejemplo la superveniencia débil, una noción intramundo, o la superveniencia natural, que sólo se mantiene a través de mundos naturalmente posibles i.e. aquellos que conservan las leyes naturales. Para una discusión en torno a la noción de superveniencia véase p. ej. Kim (1993; Cap. 4)

Vale la pena señalar también que los mundos que se consideran relevantes en la tesis del fisicalismo a través de la superveniencia no solo son aquellos que son físicamente indiscernibles, sino que además no contienen ningún agregado no-físico adicional (p.ej. fantasmas, plasma, etc). Esto es, duplicados físicos mínimos (Chalmers, 1996; 42; Jackson, 1994)

acorde a esta relación; cuando experimentamos un dolor, tomamos ácido acetilsalicílico que actúa a nivel de la química cerebral-corporal; cuando las funciones cerebrales de alguien se detienen a causa de algún golpe, decimos que el individuo en cuestión está inconsciente, por ejemplo. Así, parece que el punto de partida en la discusión es que nuestras propiedades psicológicas, incluidas aquellas que hacen a un estado mental fenoménicamente consciente, deberían tener cabida en una ontología física.

Otra forma de apoyar esta intuición es ver la ontología, el inventario de las propiedades del mundo, en distintos niveles. Así, algunas propiedades sobrevienen directamente en otras que a su vez sobrevendrían en propiedades más básicas, siendo la superveniencia una relación transitiva. Ya que las propiedades biológicas sobrevienen en las propiedades físicas y puesto que tenemos buenas razones para considerar nuestras propiedades mentales como propiedades biológicas, estas serían en última instancia propiedades físicas.

1.2.2 Dualismo

El fisicalismo es, en su sentido estricto, una ontología monista, pues sostiene que hay sólo un tipo de propiedad básica en el mundo. Negar el fisicalismo es sostener que hay por lo menos algún tipo de propiedades extra, que no sobrevienen en las propiedades físicas. Estas propiedades serían entonces tan básicas como las físicas, es decir, pudiese ser que otras propiedades sobrevengan, a su vez, en ellas. La forma más sencilla de una ontología no fisicalista sería una ontología dualista, que, como su nombre indica sostendría dos tipos de propiedades como básicas.

Dado que aquí me estoy centrando en las propiedades fenoménicas y su relación con las propiedades físicas, el candidato a ser este tipo de propiedad no física serían precisamente las propiedades fenoménicas. Si otras propiedades (como las mencionadas sociales, políticas, morales, etc...) resultan no ser ni físicas ni fenoménicas, ni metafísicamente supervenientes en ellas,

estaríamos ante una ontología pluralista más amplia. No consideraré eso aquí, por lo cual hablaré solamente del dualismo como la ontología que rechaza el fisicalismo.

La diversidad de ontologías dualistas requiere un tratamiento más amplio del que puedo dedicar aquí, pues solo presento al dualismo como la postura antagónica al fisicalismo. Por ejemplo, además de reconocer la existencia de propiedades no físicas, una ontología dualista podría sostener que existen sustancias no físicas, lo que sea que eso quiera decir. Además, el dualista tiene la misión de dar cuenta de cómo se relacionan propiedades de distinto tipo (un problema al que se enfrentan los dualistas desde Descartes, uno de los dualistas más famosos) y en esto tampoco hay una respuesta única. No diré más al respecto acerca del dualismo, pues la discusión que trato se inscribe en una ontología fisicalista, como expondré en la siguiente sección.⁸

1.3 La cuestión explicativa

Aunque también algunos dualistas estarían dispuestos a aceptar que existe algún tipo de relación íntima entre estados físicos y mentales⁹, desde el dualismo no se tiene la intención de proveer explicaciones reductivas de la conciencia en términos físicos, con lo que la discusión que aquí trato se inscribe en un marco fisicalista¹⁰. Sin embargo, se puede ser fisicalista sin suscribir un reduccionismo explicativo, por lo que de concluir que las explicaciones reductivas no son posibles,

⁸ Para una exposición detallada de la ontología dualista, sus distintas variedades, los problemas a los que se enfrenta y las soluciones de las que dispone, véase Robinson (2017)

⁹ Incluso algunos dualistas, como el mismo Chalmers (1996), pueden sostener alguna versión de la superveniencia menos fuerte que la versión metafísica, restringiendo la relación a mundos naturalmente posibles, por ejemplo.

¹⁰ Una alternativa cercana al dualismo en la que podrían buscarse explicaciones reductivas es el protopanpsiquismo. Esta ontología concibe que las propiedades fenoménicas surgen de alguna combinación de propiedades protofenoménicas, las cuales a pesar de ser propiedades intrínsecas de la materia, no son propiedades físicas en el sentido que aquí se toma como físico (aunque son, al menos, muy cercanas a la versión de “propiedad física” a partir de objetos descrita antes.) No consideraré estas posturas en este trabajo, pues por su complejidad requieren un estudio independiente. (véase Chalmers, 2015)

conclusión que pretendo apoyar, no se sigue inmediatamente la falsedad del fisicalismo. (En adelante me referiré por “reduccionismo” simplemente al proyecto de ofrecer explicaciones reduccionistas, y no a algún tipo de reduccionismo en sentido ontológico¹¹).

Es así que los proyectos reduccionistas se comprometen con distintos niveles conceptuales, como explicaré en el siguiente capítulo, pero suelen inscribirse en marcos monistas, en este caso un fisicalismo.

El reduccionismo es una tesis cuya comprobación final es empírica, pero parece haber éxito al explicar muchos fenómenos naturales (biológicos, químicos, psicológicos, etc.), en términos físicos. De tal forma, esta doctrina parece dar soporte a la tesis de la generalidad de la física y su fundamentalidad explicativa respecto a (al menos algunas de las) otras ciencias.

En ese sentido, dado que tenemos mejor acceso epistémico desde una perspectiva de tercera persona, a los estados físicos que consideramos de algún modo subyacentes a los estados mentales, vale la pena investigar si estos pueden decirnos algo, y qué exactamente, acerca de nuestra mente. Esto se ha convertido en una especie de guía metodológica de buena parte de la psicología y ciencias cognitivas contemporáneas, como ya se ha mencionado.

Sin embargo, se ha argumentado que, a diferencia de otros estados y procesos mentales, la presencia de la conciencia fenoménica se resiste a la explicación reductiva en términos físicos (Chalmers (1996), Levine (1983), Jackson (1982; 1986), McGinn (1991), Nagel (1974), entre otros). Explicar porque un estado físico posee determinado carácter fenoménico y no otro, es lo que Chalmers (1996) llama “el problema duro de la conciencia” y aproximadamente lo que Levine (1983) denomina la brecha explicativa (*the explanatory gap*).

¹¹ El reduccionismo ontológico en filosofía de la mente suele identificarse particularmente con el fisicalismo tipo (*type physicalism*). Esta postura no goza de mucha popularidad pues sostiene que cada tipo mental es idéntico a un tipo físico, con lo se enfrenta a la muy plausible tesis de la realización múltiple que explicaré en el capítulo cuarto (pág. 62).

Ya que en este capítulo he expuesto lo que es la conciencia fenoménica y he introducido brevemente lo problemática que resulta, en los siguientes capítulos expondré precisamente la cuestión de la explicación reductiva, para lo cual comenzaré por limitar la discusión a un modelo de explicación reductiva en el siguiente capítulo. En el tercero, por otro lado, expondré los argumentos que nos llevan a concluir que la conciencia no puede ser explicada reductivamente.

Capítulo 2:

Explicaciones reductivas y análisis conceptual

El propósito de esta tesis es evaluar la posibilidad de explicaciones reductivas de la conciencia fenoménica. Comencé esta discusión aclarando lo que es la conciencia fenoménica. En este capítulo expondré lo que será considerado como una explicación reductiva en la sección 2.1. El modelo explicativo que tomaré como central será el nomológico deductivo (ND). Sostendré que una explicación de este tipo cumple con ser una explicación reductiva al cumplir un requisito adicional.

Expondré también un problema de tensión aparente entre el modelo expuesto y el requisito solicitado y una solución que apela a la semántica de los términos en que se describe aquello que quiere explicarse. Para esto y para la discusión siguiente será particularmente útil relacionar este modelo con la tesis de la implicación *a priori*, esto lo haré en la sección 2.2. Además, en las secciones 2.3 y 2.4 mostraré que las explicaciones reductivas operan mediante consideraciones semánticas. De esta forma, al final del capítulo se verá qué se considerará una explicación reductiva, qué forma tienen estas y cómo pueden ser satisfechas.

2.1 Explicaciones reductivas

Como ya se ha expresado, el tema de esta tesis es el de las explicaciones reductivas de la conciencia fenoménica. Dedicaré este capítulo a explicar lo que consideraré una explicación. Aunque dentro del debate en torno a la conciencia se discute si existe algo así como una brecha explicativa, o si la conciencia puede explicarse reductivamente, etc., pocas veces se hace explícito de qué forma se entiende que es una explicación. Ésta, considero, no es una cuestión menor, pues la conclusión podría variar relevantemente dependiendo de lo que se tome como tal. Así, en este capítulo centraré el debate a un modelo de explicación científica particular: el modelo nomológico deductivo.

Hay dos razones para limitar la discusión a este modelo. La primera es que considero es este el modelo supuesto en gran parte del debate, o por lo menos es muy compatible con él. La segunda es que este modelo de explicación es particularmente relevante pues ha sido una de las primeras propuestas de lo que es una explicación científica y a pesar de las muchas críticas que ha recibido, no parece estar del todo abandonado.

Cotidianamente entendemos por *explicar* el dar cuenta de la ocurrencia de un fenómeno, es decir, responder a la pregunta de el *porqué* de su ocurrencia. Explicar es, en cierto sentido, hacer inteligible un suceso. Por ejemplo, explicar el hecho de que me encuentro escribiendo una tesis, es, aproximadamente, contar que lo hago porque deseo titularme y sé que esto es un requisito necesario para lograrlo; explicar el hecho de que mi taza se rompiera al caer es mostrar que esto es un suceso esperable dadas ciertas características del material con que fue hecha y las condiciones del golpe que sufrió con la caída, etc.

Es usual pensar que una de las tareas de la ciencia es precisamente explicar algunos sucesos, pero a diferencia de una explicación ordinaria, tendrá que hacerlo subsumiéndolos a leyes y teorías científicas. Por ejemplo, explicar la órbita de los planetas puede hacerse mostrando que la manera en que se mueven está en conformidad con las leyes de Newton (exceptuando el famoso caso de la

anomalía en el perihelio de Mercurio, que, sin embargo, puede explicarse con leyes de la teoría general de la relatividad); o bien explicar el evaporarse del agua al hervir puede hacerse a partir de mostrar como esto es esperable de acuerdo a como el calor suministra la suficiente energía para romper la tensión superficial del líquido, dadas ciertas condiciones de presión atmosférica, y como esto es esperable de acuerdo a ciertas leyes. En este sentido explicar es una actividad distinta y posterior a la formulación y/o descubrimiento de estas leyes, teorías o hipótesis.

A mediados del siglo pasado, Hempel y Oppenheim (1948) propusieron un modelo de explicación científica que rescata gran parte de esto, y que ha sido el punto de partida de la discusión en torno a este tema: el modelo nomológico deductivo (ND). En esta tesis no pretendo sumarme al debate acerca de la total adecuación de este modelo, sino asumiendo que bajo él se pueden proveer explicaciones de algunos fenómenos, evaluaré si la ocurrencia de la conciencia fenoménica puede explicarse reductivamente en esos términos¹².

Según Hempel y Oppenheim, una explicación científica no es más que una deducción en la cual se concluye la descripción del fenómeno que se busca explicar (lo que denominan *explanandum*) partiendo, precisamente, de leyes científicas (*L*) y determinadas condiciones iniciales (*C*), que juntas forman lo que denominan *explanans*. De tal forma, un argumento-explicación según el modelo ND tendría la siguiente estructura general (Hempel y Oppenheim, 1948; 138):

$$\begin{array}{r}
 C1 \dots Cn \\
 L1 \dots Ln
 \end{array}
 \left. \vphantom{\begin{array}{r} C1 \dots Cn \\ L1 \dots Ln \end{array}} \right\} \textit{Explanans}$$

$$E \left. \vphantom{E} \right\} \textit{Explanandum}^{13}$$

¹²Algunos argumentos en contra del modelo, en particular respecto a su suficiencia, pueden encontrarse recopilados en Strevens, M (2006, 4) y Salmon (1989, Cap. I)

¹³ De acuerdo con estos autores, cuando el fenómeno descrito en el *explanandum* no ha ocurrido, el modelo ND se presenta como un modelo de predicción científica. Explicar es también, en este sentido, mostrar que la ocurrencia de un fenómeno es esperable a partir de determinadas leyes y condiciones.

Considerar una explicación sólo como un argumento deductivo se enfrenta con varias dificultades formales, pero podemos evadirlas apelando a criterios y requisitos adicionales. La primera es que cualquier cosa puede seguirse de sí misma, por lo que E explicaría a E , sin embargo, podemos considerar una deducción de este tipo como trivial y nada informativa, rechazándola como una explicación¹⁴.

La segunda dificultad es que de tener premisas falsas o contradictorias, cualquier cosa podría concluirse con validez, con lo que, de acuerdo con el modelo, podríamos explicar cualquier cosa con premisas de este tipo. Pero podemos simplemente estipular que para que un argumento sea una buena explicación las premisas deben ser todas verdaderas (Hempel y Oppenheim, 1948; 137).

La última dificultad es la más significativa. En la conclusión de un argumento deductivo cuyas premisas son todas verdaderas, tal como lo solicitamos, no puede aparecer una premisa nueva.¹⁵ E debe, entonces, estar presente dentro del *explanans*, ya sea dentro de las condiciones iniciales o como parte de las leyes. El primer caso trivializaría la explicación pues, haría innecesaria la inclusión de leyes científicas, obteniendo un argumento de la forma $E \vdash E$, que ya he señalado como irrelevante. Queda pues que la premisa E forme parte de alguna ley científica.

Quienes suscriben el modelo ND formulan los enunciados legaliformes como condicionales universales¹⁶. Una ley, entonces, debe tener la forma $C \rightarrow E$, lo que sumado a determinada condición C , permite concluir E :

¹⁴ La misma irrelevancia ocurriría si se tuviese algo del tipo $E \& I \vdash E$

¹⁵ A menos que ésta premisa nueva sea un disyunto de una premisa que sí esté presente. Por ejemplo, de la premisa I se sigue válidamente que $(E \rightarrow I)$, $(E \vee I)$, etc. pero no se puede seguir simplemente E . Estos casos pueden descartarse, pues no obtenemos una conexión significativa entre *explanans* y *explanandum* para contarlos como explicaciones de E .

¹⁶ Dado que la noción de ley científica es central en el modelo ND, esta ha sido uno de los puntos más discutidos del mismo y la noción merece algunas aclaraciones. Para Hempel y Oppenheim (1948; 152) las leyes científicas son generalizaciones universales, expresables en forma condicional (aunque no toda generalización de este tipo es una ley (por ejemplo "Todos los que están en este salón son adictos al café") el debate en torno a lo que las distingue es externo a este trabajo (Cf. Salmon, 1989; 14), por lo que apelaré a nociones más o

$$\frac{C \quad C \rightarrow E}{E}$$

Ahora bien, ¿Cuándo una explicación de acuerdo a este modelo es reductiva? Como se adelantó en la última sección del capítulo anterior, la reflexión en torno a las explicaciones reductivas está comprometida con la existencia de distintos niveles. Esta jerarquía de niveles no es ontológica, de hecho es mejor considerarla como una jerarquía puramente conceptual, basada en los términos que se usan en las descripciones de los fenómenos. Por ejemplo, los términos en el nivel científico, con la física a la base, son conceptualmente distintos de los de un nivel más o menos cotidiano aunque términos como “NaCl”, “cloruro de sodio” o “sal” refieran todos al mismo ente.

Buena parte de la jerarquía de niveles está dada precisamente porque algunas leyes pueden explicar otras, resultando así que algunas ciencias, y sus términos, son más básicas. En el modelo ND, E no es necesariamente la descripción de un fenómeno particular, puede ser también de alguna regularidad o ley (Hempel y Oppenheim, 1948; 137). Tenemos buenas razones para considerar que las leyes de la física son más básicas pues en sus términos se pueden deducir muchas de las leyes de la química, la biología y otras ciencias naturales.¹⁷ De hecho, como ya se dijo, la discusión que aquí trato se inscribe dentro del fisicalismo, que considera las propiedades físicas como fundamentales.

Así, una explicación reductiva tendría que explicar un fenómeno o regularidad en términos de un nivel considerado inferior. Explicar un fenómeno en términos de otros del mismo nivel no

menos intuitivas). De tal forma, una ley puede expresarse más finamente por medio de cuantificadores como $\forall x (Cx \rightarrow Ex)$.

¹⁷ El ejemplo ofrecido por Hempel y Oppenheim son las leyes de Kepler que pueden explicarse apelando a las leyes de Newton, que a su vez pueden deducirse, finalmente, de la teoría de la relatividad general (Hempel y Oppenheim, 1948; 146)

tendría mucho de “reductivo”. En otras palabras, en una explicación reductiva, los términos en que se presenta el *explanandum* deben ser de un nivel distinto, e inferior, a los del *explanans*.

Así, por ejemplo, explicar reductivamente un fenómeno mental, es, a grandes rasgos, dar cuenta de su ocurrencia a partir de fenómenos que consideramos de un nivel distinto y fundamental, paradigmáticamente físicos. Por el contrario, explicar un suceso de este tipo en términos de otros sucesos mentales, no se consideraría una explicación reductiva. Tampoco lo sería de no hacerse el paso de los términos microfísicos a los términos en los que originalmente se planteó el *explanandum*, en este caso a términos mentales.

Esto es precisamente lo que da valor epistémico a una explicación reductiva. Aunque no siempre buscamos explicaciones de este tipo, tanto en ciencia como en la vida cotidiana estas son útiles porque nos ayudan a hacer inteligible algún fenómeno partiendo de otros que conocemos de mejor manera y que consideramos tienen una relación de determinación o incluso de identidad con aquello que queremos explicar.

Kim, un filósofo de la mente que se ha dedicado también a las explicaciones científicas, formula explícitamente este requisito como: “(R): Las premisas *explicativas* de una explicación reductiva de un fenómeno en el que está presente la propiedad *F* (p. ej. Una explicación de porqué *F* es instanciada en determinada ocasión) no deben referir a *F*” (Kim, 2005; 106. Traducción y énfasis míos), o bien, en una formulación más exigente:

(R): En el *explanans* de una explicación reductiva de *F* no sólo no se debe referir a *F*, sino tampoco debe referirse a cualquier otra propiedad del nivel de *F* –o, equivalentemente, una explicación reductiva de la ocurrencia de *F* debe referir en su *explanans* sólo a propiedades de nivel inferior a *F* (Kim, 2005; 106. Traducción mía).

A lo largo de este trabajo me referiré nuevamente a este requisito a través de esta misma formulación como “(R)”.

A primera vista existe un problema con los modelos de explicación deductivos y la intención de establecer explicaciones reductivas de acuerdo a este requisito. Si para tener una explicación reductiva requerimos que los términos en los que E se presenta no estén incluidos explícitamente en las premisas explicativas, esto resulta problemático si una explicación es un argumento deductivo, pues E tendría que aparecer forzosamente en el *explanans* o de otra forma no existiría ninguna conexión inmediata entre las premisas y la conclusión de la explicación.¹⁸

Es necesario, entonces, algo que permita la traducción entre los términos usados en los distintos niveles, pero ¿se puede ofrecer una premisa de este tipo sin violar la condición (R)? Para presentar la solución a este problema, ligaré primero el modelo presentado con una tesis que en gran medida se discute ya en el debate en torno a la conciencia: la tesis de la implicación *a priori*. Expondré ésta en la siguiente sección.

2.2 La tesis de la implicación *a priori*

Ya en el contexto de las discusiones en torno a las explicaciones de la conciencia, Chalmers y Jackson (2001) proponen la tesis de la implicación *a priori* como una condición necesaria para las explicaciones reductivas, sin ligarlas explícitamente a un modelo particular. Después de explicar esta tesis y relacionarla con el modelo de explicación ofrecido, mostraré como desde ella se ofrece una solución al problema planteado al final de la sección anterior.

Según la tesis de la implicación *a priori*¹⁹ (TIA), formulada por Chalmers y Jackson (2001; 316) el condicional $P \rightarrow M$, en donde P es una descripción hipotética completa del mundo en los

¹⁸ Kim (2005; 98) reconoce también este problema y lo llama el problema del ascenso explicativo (*the explanatory ascent problem*). Su solución es similar a la que expondré más adelante.

¹⁹ En Chalmers y Jackson (2001) la formulación es explícita y más desarrollada. Las ideas en torno a la implicación *a priori* no nacieron precisamente en ella. Véase, por ejemplo Chalmers (1996) o Jackson (1993: 1998 cap. 2 y 3), en donde están presentes.

términos microfísicos más fundamentales, esto es, en aquellos que refieren directamente a las propiedades físicas básicas; y M la descripción de un fenómeno natural cualquiera formulada en términos macroscópicos, debe conocerse *a priori* para tener una explicación reductiva de M en los términos de P .

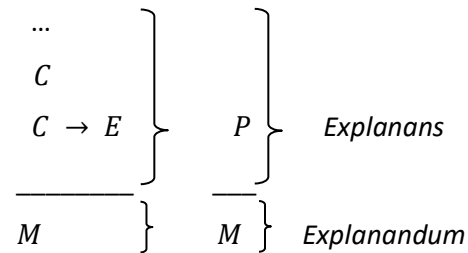
Pero además, estos autores agregan al antecedente del condicional, información índice I , como “ahora”, “aquí” o “yo”; una cláusula T (*that is all*) que excluye la implicación de hechos negativos y de aquellos que no se siguen estrictamente de P ²⁰; y una que incluye los hechos fenoménicos Q . De tal forma, es necesario que el condicional resultante $(P \wedge Q \wedge T \wedge I) \rightarrow M$ se conozca *a priori* para ofrecer una explicación reductiva de M . Como el propósito es justamente evaluar si los hechos fenoménicos son implicados, no incluiré Q en el antecedente sino en el consecuente. En cambio, por practicidad, me referiré al antecedente simplemente como P , incluyendo allí también T e I . (Chalmers y Jackson, 2001; 317-319).

Chalmers y Jackson dan sólo algunas motivaciones e intuiciones para inclinar la balanza a favor de esto. Si sostenemos una ontología fisicalista, los hechos descritos en M están de hecho ya incluidos en P , aunque en un vocabulario distinto (Chalmers y Jackson, 2001; 330-333). Se puede comparar que el condicional de la tesis de la implicación *a priori* es análogo al de la tesis de la entrada por implicación, por la que se define el fisicalismo, con la diferencia de que en este último la implicación es de carácter metafísicamente necesario. Así, la TIA es plausible, aunque no por ello garantizada, dentro de una ontología fisicalista.

Si bien, estos autores no se comprometen explícitamente con el modelo ND, sospecho que es el modelo de explicación que de hecho está supuesto. Mi propósito en lo que sigue es mostrar que el modelo ND y La TIA son compatibles y pueden resultar complementarios. (Cf. Carruthers,

²⁰ Esta cláusula juega un papel análogo a la restricción de duplicados físicos que no contienen nada más al formular el fisicalismo.

2005; 21-27; en donde también se expone la relación de la tesis de la implicación *a priori* con el modelo ND y otros modelos de explicación, favoreciendo igualmente su compatibilidad con el primero). De esta forma, al establecer un puente estructural entre el modelo ND y la TIA, obtendríamos el siguiente esquema:



Puesto que P es una descripción completa del mundo, $C \rightarrow E$ y C claramente no se identifican con P , pero pertenecen a ella, el *explanans* mismo es entonces un subconjunto de P . Chalmers y Jackson consideran P , la descripción completa, para evitar los casos en los que no podemos explicar reductivamente a M por falta de información, pues *ex hyphotesi*, tenemos toda los datos posibles. Esto será relevante al considerar los argumentos en contra de la explicación reductiva de la conciencia en el segundo capítulo.

Por otro lado, el *explanandum*, bien puede no ser una descripción en términos macroscópicos, pero, si la explicación se plantea reductiva este debe ser una descripción en términos de un nivel más alto que los usados en el *explanans*, tal como Chalmers y Jackson proponen que es M respecto a P . Acorde al requisito formulado, los términos usados en M entonces no pueden estar presentes explícitamente en alguna de las premisas explicativas que componen P (y por tanto tampoco en el *explanans*), tal como muestra el esquema anterior.

En el esquema anterior, se ve claramente el problema de tensión entre (R) y el modelo ND que mencioné al final de la última sección anterior, es decir, no es claro cómo deducir el *explanandum* de premisas en donde los términos de éste no aparecen.

No obstante, aquello que de acuerdo con Chalmers y Jackson permite la a prioridad entre P y M es justamente lo que permite el paso entre *explanans* y *explanandum* en una explicación reductiva, permitiendo conservar el requisito (R): cierto tipo de análisis conceptual. Lo que se concluye en el argumento es la ocurrencia de un fenómeno, que si bien está inmediatamente descrito en los términos en los que el mismo se presenta en las premisas, es, en principio, traducible a términos distintos, aquellos del *explanandum*, mediante un análisis conceptual del significado que puede ser realizado *a priori* por quien sea que posea los conceptos usados en M . Esto es, lo que permite la satisfacción de la explicación reductiva son consideraciones semánticas de los términos involucrados.

Dado que este análisis conceptual es *a priori*, de ser satisfecho no se necesitaría incorporar más premisas empíricas en el *explanans* para poder realizar el paso entre los distintos vocabularios, con lo que se satisfaría el requisito (R); en la otra dirección, de no poder cumplirse la tesis de la implicación *a priori*, significaría que una premisa empírica extra con contenido empírico tendría que agregarse al *explanans*. Antes de profundizar y ejemplificar esta relación entre la TIA y (R), expondré en la siguiente sección como es el análisis conceptual que permite la explicación reductiva.

2.3 Análisis conceptual y significado

Para Chalmers y Jackson el significado de un concepto está dado en gran medida por su intensión. Una intensión es una función $f: W \rightarrow R$, que toma como *inputs* mundos posibles W y arroja como resultado extensiones R .²¹ (Chalmers, 1996; 89). Es decir, el significado de un concepto

²¹ Brevemente la extensión de un término simple es el objeto al que refiere. Para el caso de proposiciones es su valor de verdad y para el caso de términos generales, como clases naturales, es el conjunto de objetos que caen dentro del término (por ejemplo, la extensión de “perro” es el conjunto de todos los perros), etc.

no es (o no es solamente) su referencia, sino la función por la que la “selecciona”, dependiendo de cómo resulta ser el mundo.²²

Estas funciones, en tanto mecanismo de asignación del referente, son realizadas por quienes conocen el significado de un término, brevemente, por quienes pueden usarlo eficazmente en prácticas lingüísticas e inferenciales. Esto es a grandes rasgos lo que significa poseer un concepto. Si bien la intensión de un término no necesita expresarse explícitamente, se puede dar un análisis aproximado de la misma mediante una descripción en términos funcionales de los roles que satisface el referente del término²³ (Chalmers, 2002; 148-149). Por ejemplo, para el término “agua”, la intensión puede ser expresada aproximadamente por una descripción del tipo “el líquido que llena los ríos y los lagos, que cae del cielo, que es esencial para la vida, es incoloro, inodoro e insaboro,...”. Esta función, en el mundo real, seleccionará como referente al compuesto químico H₂O, pues este es quien satisface estos roles (o gran parte de ellos).²⁴

Vale la pena explicitar que no es necesario que la intensión tome como argumento el mundo real, pueden ser usados escenarios posibles. Dependiendo de cómo resulten ser éstos el resultado de la función será distinto, pero teniendo una descripción relevante y suficiente de este escenario hipotético, el sujeto que conoce un término es capaz de asignar un referente en dichos escenarios. Por ejemplo, en la tierra gemela de Putnam (1975), en donde el líquido que llena los ríos, mares y

²² Aunque en Chalmers y Jackson (2001; (322-326)) no se habla explícitamente de “intensión”, como sí en Chalmers (1996) o Jackson (1998), sino de un condicional de evaluación de referentes, considero que es esencialmente la misma noción, aunque la función “intensión” me parece más detallada y comprensible.

²³ De acuerdo con Chalmers (1996, 91), lo que aquí se sostiene no es necesariamente un descriptivismo, pues no se dice que los nombres sean semánticamente idénticos a las descripciones. Aquí, las descripciones solamente ayudan a expresar la función por la que se fija el referente de algunos términos y para cada término y sujeto que lo conoce, puede existir más de una. De hecho, para Chalmers esto es incluso compatible con una teoría causal de la referencia.

²⁴ En filosofía del lenguaje son usuales los ejemplos basados en los términos “agua”, “H₂O” y su referente, el compuesto H₂O. He decidido usarlos también, pero es fácil pensar en otros términos de clases naturales.

lagos, que cae del cielo, etc. no es H_2O , sino XYZ, el compuesto que seleccionaría la intensión de “agua” sería precisamente XYZ.²⁵

Aunque en el mundo actual dos términos pueden tener el mismo referente, estos pueden diferir respecto a sus intensiones. Al tratar con distintos escenarios posibles estas intensiones podrían llevar a extensiones distintas, tal es el caso de “Agua” y “ H_2O ”: la de “agua” es la que se ha descrito antes, mientras que la de “ H_2O ” está estrechamente vinculada a nociones científicas-teóricas, “el compuesto resultante de la mezcla de dos moléculas de hidrogeno y una de oxígeno”, por ejemplo. Aunque ambas seleccionan el mismo compuesto en el mundo actual, no ocurriría lo mismo en la tierra gemela.

“Héspero” y “Fósforo”, por poner un famoso ejemplo Fregeano, refieren ambos al mismo objeto: al planeta Venus. Sin embargo, los modos de selección del referente de estos términos, la estrella más luminosa de la tarde y la estrella más luminosa de la mañana, pueden seleccionar objetos distintos en otros escenarios posibles, aquellos en los que Venus no es la estrella más luminosa de la mañana o de la tarde. De aquí que se señale exista una diferencia conceptual entre los términos “Héspero” y “Fósforo”.²⁶

²⁵ No es necesaria una discusión metafísica profunda acerca de qué tipo de posibilidad se está considerando, basta señalar que se trata de una noción de posibilidad epistémica. Un escenario es epistémicamente posible si no puede ser rechazado a partir de lo que se conoce. Esta noción es sujeto dependiente, pero podemos obtener una noción filosóficamente más útil si postulamos un sujeto ideal carente de conocimiento empírico. Así, una posibilidad epistémica, en el sentido relevante, es aquella que simplemente no puede ser rechazada *a priori*. De acuerdo a la noción sujeto-dependiente, alguien que sabe que agua es H_2O no consideraría posible que agua sea XYZ; por el contrario, careciendo de ese conocimiento empírico, no se puede rechazar *a priori* la posibilidad de que agua sea igual a XYZ, ABC u otro compuesto cualquiera. Párrafos adelante diré algo acerca de la relación entre estos escenarios epistémicamente posibles y los mundos metafísicamente posibles.

²⁶ De hecho, es esta diferencia la que da cuenta de que identidades como Héspero=Fósforo o Agua= H_2O resulten informativas y epistémicamente relevantes. Si el significado de ambos términos en la identidad se agotara en sus extensiones, estas identidades tendrían el mismo peso cognitivo que una identidad de la forma “a=a”. En consecuencia, no sería posible que alguien creyera algo acerca de Fósforo pero lo contrario acerca de Héspero, o viceversa, lo cual es evidentemente posible para quien no sabe que Héspero es Fósforo. Esto es lo que ha llevado a algunos autores (p.ej. Chalmers, 2002) a sostener las intensiones como una especie de reivindicación del sentido Fregeano (aunque no son exactamente nociones idénticas), ya que rescatan mucho de lo que Frege (1892 (2016)) pretendió al postular los sentidos.

Existe, sin embargo, un sentido, señalado por Kripke (1980 (2017)), en el que es imposible que Héspero y Fósforo sean objetos distintos. Vale la pena detenerme un poco en aclarar esto para evitar algunas confusiones. Para esto expondré el llamado marco bidimensionalista del significado, particularmente la versión defendida por Chalmers (p. ej. 2010), desde el cual pueden rescatarse tanto la tesis kripkeana como la noción de intensión explicada.

2.3.1. Bidimensionalismo

De acuerdo con Kripke (1980 (2017; 51)), los nombres propios y términos de clase natural designan rígidamente, lo que quiere decir que tienen el mismo referente en cualquier escenario posible en el que éste existe y no refieren a nada en los cuales no. En consecuencia, la identidad entre Héspero y Fósforo, o entre agua y H_2O , es necesaria. Es decir, no es posible que Héspero no sea Fósforo (pues ambos son Venus) o que agua sea algo distinto a H_2O . A grandes rasgos, podemos pensar en el papel de los designadores rígidamente en la teoría de Kripke de forma análoga a lo que en las teorías bidimensionalistas del significado cumplen las llamadas intensiones secundarias.

Una teoría bidimensionalista postula la existencia de dos intensiones para cada término (Chalmers, 2010): las intensiones primarias, o epistémicas, y las intensiones secundarias, o metafísicas. Las intensiones primarias son las que ya he explicado; las intensiones secundarias, por otra parte, también son también funciones de mundos posibles a referentes, pero difieren en la forma en que consideran los mundos que toma como entradas. Mientras que las intensiones primarias seleccionan el referente considerando un escenario posible como actual, las intensiones secundarias lo hacen considerándolo como contrafáctico, esto es, seleccionan el referente en escenarios posibles a partir de como lo hacen en el mundo que se considera como actual. De esta forma la intensión secundaria queda mejor descrita como la función $f: (W^* \times W) \rightarrow R$, en donde

W^* es un mundo que se considera como actual, W un mundo contrafáctico desde W^* y R un referente (Chalmers, 1996; 94).

Por ejemplo, la intensión secundaria del término “agua” partiendo del mundo actual nos dice que este término seleccionará el mismo referente en los que el compuesto H_2O exista y nada en los que no. De acuerdo con esto, el líquido XYZ de la tierra gemela no sería agua, sino sólo una sustancia superficialmente similar. Análogamente, si partimos de considerar la tierra gemela como actual y concedemos que “agua” refiere a XYZ, un mundo en el que el líquido que llena los ríos y lagos,... etc. resulta ser H_2O , sería un mundo en el que no habría agua pero sí una sustancia similar a ella.

De esta forma, obtenemos una matriz de los referentes de un término según se consideren las intensiones primaria y secundaria del mismo. Siguiendo con el ejemplo del término “agua”, por ejemplo (Chalmers, 2010; 543):

	a) Mundo H_2O	b) Mundo XYZ	c) Mundo...ABC
1) Mundo H_2O	H_2O	H_2O	H_2O
2) Mundo XYZ	XYZ	XYZ	XYZ
3) Mundo ABC	ABC	ABC	ABC

Si partimos de considerar los mundos de la fila de arriba (a, b, c...) como actuales y nos movemos hacia abajo, considerando 1), 2) 3)... como escenarios posibles, obtenemos un referente distinto para cada intensión: partiendo de a) como actual, obtenemos H_2O para a1, XYZ para a2...; este movimiento representa a la intensión primaria. En contraste si consideramos los mundos de la fila izquierda (1, 2...) como actuales y nos movemos hacia la derecha, obtenemos el mismo referente en mundos contrafácticos: partiendo de 1) como actual, obtenemos H_2O en 1a, H_2O en 1b,..., etc. esto corresponde a la intensión secundaria.

Así, es de acuerdo a la intensión secundaria, a partir del mundo actual, que es imposible que agua sea XYZ dado que ésta es necesariamente H₂O, rescatando el sentido de necesidad señalado por Kripke. A la vez, mediante la intensión primaria se puede dar cuenta de la apariencia de contingencia en las identidades de este tipo, es decir, explican el sentido en el que parece posible que Héspero no sea Fósforo o que agua no sea H₂O a pesar de que estas sean necesidades.

De acuerdo con Kripke, estos escenarios epistémicos se corresponderían con mundos metafísicamente posibles, pero serían mundos mal descritos. Por ejemplo, en los mundos en los que hay XYZ y no H₂O, lo que hay no es agua, sino una substancia superficialmente similar que denominamos e identificamos erróneamente con “agua”. Así, dado que esta contingencia se basa en nuestro acceso epistémico a los escenarios posibles, se rescata la distinción conceptual entre “agua” y “H₂O” o entre “Héspero” y “Fósforo”.

Es la intensión primaria la que es relevante para cuestiones epistémicas puesto que está directamente relacionada con lo que un sujeto podría saber dado cierto escenario hipotético que se considere como actual y la posesión de un concepto. Además, evaluar escenarios como posibles no depende de cómo resulte ser el mundo que se tome como real, por lo cual la función de la intensión primaria es *a priori*. Es esta intensión la que me interesa aquí.

La intensión secundaria suele estar asociada a debates de corte metafísico, pues al evaluar mundos contrafácticos depende totalmente de cómo resulta ser el mundo que se considera como actual, con lo cual es *a posteriori* pues esta información sólo se conoce empíricamente. Es por ello que se conoce a las intensiones primaria y secundaria como intensiones epistémica y metafísica, respectivamente.

2.4 Explicación reductiva y análisis conceptual

Así, Para Chalmers y Jackson, quien posee el concepto “agua” es capaz de realizar un análisis del significado del mismo en términos funcionales (mediante la intensión primaria), y dada una descripción suficiente y relevante de algún escenario epistémico hipotético, en este caso expresado por P , de fijar como referente a aquellos elementos que satisfacen la función. Esto es a lo que se refieren por “análisis conceptual”. Para esto no se requiere información empírica adicional, pues en posesión de P se sabe ya qué cumple esos roles, resultando en la *a prioridad* entre P y M .

Volviendo a nuestro modelo explicativo, lo que estrictamente se concluye en la deducción es una descripción en términos microfísicos, efectivamente ya incluidos en P , de algún suceso. El paso posterior para hacer de esta una explicación reductiva es conectar esta descripción con los términos macroscópicos adecuados en los que se presentó originalmente la situación a explicar. Este segundo paso puede realizarse sin necesidad de información empírica adicional a P por medio del análisis conceptual mencionado. Este paso no siempre es explícito en el argumento, pues puede ser realizado por los agentes epistémicos que se encuentran en la situación adecuada. (Chalmers y Jackson, 2001; 332-333). Por ejemplo:

$$\begin{array}{l}
 \left. \begin{array}{l}
 \dots \\
 p_n. \text{ las moléculas de NaCl se comporta de tal y tal manera en H}_2\text{O} \\
 (p_{n1}. \text{ “Agua”} =_{\text{def}} \text{ El líquido que } \dots) \\
 (p_{n2}. \text{ “Sal”} =_{\text{def}} \text{ Los cristales que } \dots) \\
 p_{n3}. \text{ H}_2\text{O es el compuesto que en el mundo actual que satisface } p_{n1} \\
 p_{n4}. \text{ NaCl es compuesto que en el mundo actual que satisface } p_{n2} \quad \dots
 \end{array} \right\} P \\
 \hline
 M \left\{ \begin{array}{l}
 \therefore \text{ La sal se comporta de tal y tal manera en el agua}
 \end{array} \right.
 \end{array}$$

Aunque las premisas p_{n1} y p_{n2} se señalan como definiciones, no lo son en sentido estricto, sino que son las premisas que expresa el análisis conceptual referido (que podría no ser una definición estricta en términos de suficiencia y necesidad). Como ya se ha dicho, estas premisas son

en realidad un paso epistémico realizado por el sujeto que posee el concepto “agua” y la información pertinente del escenario *P*. Más importante aún, su presencia en el *explanans* no viola el requisito (R), pues la ocurrencia de los términos “agua”, “sal” (o cualquiera que sea el término), no refiere directamente a un ente, sino que son premisas conceptuales, es decir, no tienen un contenido empírico. (En los términos de Kim, estas no son premisas explicativas).

Así, cuando se cumple la tesis de la implicación *a priori*, se satisface (R); en el cuarto capítulo mostraré que al no cumplirse, tampoco (R) es satisfecha. Pero además, se puede relacionar de mejor manera el requisito (R) en el modelo ND con la TIA, ya que al exponer esta tesis, Chalmers y Jackson hacen énfasis en la deducibilidad, pero esta no es meramente lógica, sino que involucra factores semánticos.

De acuerdo con esto, algunos autores señalan (por ejemplo por Chalmers, 1996; Kim, 2005) que una explicación reductiva es un proceso de dos partes. Una, el análisis conceptual; la otra, mostrar qué propiedades satisfacen el análisis del concepto ofrecido. Si bien el segundo puede ser un paso empírico, pues en el mundo real es conocer *P* (o sus aspectos relevantes), la *a prioridad* recae en la justificación de la deducción de la forma en que se explicó.

Existen algunas complicaciones menores, pero merece la pena aclarar desde ahora (Estas aclaraciones se encuentran también en el texto de Chalmers y Jackson (2001; 336-350), ya que este es en gran parte una respuesta a las críticas de Block y Stalnaker (1999) quienes formulan las objeciones presentadas, y una propuesta de la que hablaré en el cuarto capítulo).

En algún escenario puede haber dos o más compuestos que satisfagan la función asociada al término “agua”, por ejemplo, puede haber XYZ en los océanos y H₂O en los mares y lagos. Si esto es así, por ejemplo, de “H₂O cubre el 70% del planeta” no podría seguirse que “agua cubre el 70% del planeta”, pues pudiese cubrir mayor superficie si consideramos que XYZ también es agua. Esta complicación es solo aparente pues en un mundo tal, la descripción completa del mundo de la que

se parte incluiría también la información sobre la superficie cubierta por XYZ y por lo tanto también sabríamos qué porcentaje del planeta cubre XYZ. (Chalmers y Jackson, 2001; 340)

Por otro lado, los mundos en los que pudiese existir una hipotética sustancia no física que cumple con los roles del agua, que en la literatura se denomina convenientemente como “*ghost-water*”, cuya descripción no está incluida en *P*, son evitados por la cláusula *T*. Estos casos entonces no son tampoco problemáticos. ((Chalmers y Jackson, 2001; 341))

Finalmente, podemos señalar que, la propia identidad entre agua y H₂O es deducible *a priori* de *P*, pues se sabe que los roles asociados al agua son satisfechos *únicamente* por el compuesto físico H₂O en el mundo actual, pues no hay otro compuesto físico en *P* que los satisfaga y no hay tampoco un compuesto no físico que lo haga, por la cláusula *T*. Si bien es cierto para conocer identidades como esta hace falta conocimiento empírico, en un escenario hipotético en donde poseemos todo el conocimiento empírico físico, la identidad se sigue de eso con *a prioridad*. Cuando la identidad “agua= H₂O” es implicada *a priori* por *P* (en el mundo actual), diremos que no es epistémicamente primitiva, pues puede ser deducida de premisas más básicas²⁷ (Chalmers y Jackson, 2001; 354). Esto será importante en el último capítulo, al considerar el posible papel de las identidades en las explicaciones reductivas.

En este capítulo he relacionado la tesis de la implicación *a priori*, central en el debate en torno a las explicaciones reductivas de la conciencia, con un modelo de explicación científica particular, el modelo ND. Además, he descrito y ejemplificado la forma que tienen las explicaciones reductivas de acuerdo con este modelo. La tarea en el siguiente capítulo será evaluar si podemos proveer explicaciones reductivas de la conciencia fenoménica, inclinándome por una respuesta negativa. Mientras que en el capítulo final, evaluaré un par alternativas de explicación reductiva,

²⁷ No debe confundirse el que la identidad sea implicada *a priori* por *P*, con que la identidad sea en si misma *a priori*, de hecho “agua=H₂O” es *a posteriori* en el sentido de que tiene la forma *a=b*.

aun de acuerdo con el modelo ND, desde las cuales se ha propuesto que la conciencia puede explicarse reductivamente. Sin embargo, concluiré que, aunque esta es efectivamente la forma en la que podemos dar explicaciones de la conciencia, estas no son reductivas pues incumplen con la tesis de la implicación *a priori*, y con ello con (R), y al hacerlo tienen desventajas epistémicas frente al modelo ofrecido hasta ahora.

Capítulo 3:

Las explicaciones reductivas de la conciencia fenoménica

En el capítulo anterior expuse que para que una explicación reductiva de M en términos de P pueda ser satisfecha es necesario que $P \rightarrow M$ sea cognoscible *a priori*. Además, expliqué que es gracias al análisis conceptual de los términos de M que este condicional puede ser satisfecho. En otras palabras, para explicar reductivamente un suceso debe proveerse un análisis conceptual de los términos de acuerdo a sus roles funcionales-causales, a partir de sus intensiones, que permita mostrar como los referentes de las descripciones en términos microfísicos son quienes cumplen esos roles.

Para mostrar que las explicaciones reductivas no son posibles, explicaré que el condicional $P \rightarrow Q$, en donde Q es la descripción de algún suceso fenoménico específico que busca explicarse, no es *a priori*, es decir, la tesis de la implicación *a priori*, necesaria para la explicación reductiva, no se cumple. Esto en gran medida porque no podemos proveer del análisis conceptual requerido de los conceptos fenoménicos. Para ello me serviré de un par de argumentos que se formulan directamente contra el condicional $P \rightarrow Q$: el argumento de la concebibilidad (Chalmers, 1996), que expondré en la siguiente sección, y el argumento del conocimiento (Jackson (1982; 1986) y Nagel (1974)), que presentaré en la sección 3.2. Finalmente, expondré, en la sección 3.3, que estos argumentos descansan en una distinción conceptual entre conceptos fenoménicos y psicológicos.

3.1 El argumento de la concebibilidad- Zombies

El argumento de la concebibilidad trata de reforzar una distinción conceptual entre estados físicos y fenoménicos, aún en el caso en que estados como dolor y estimulación de fibras C (o cualquiera que resulte ser su correlato en el cerebro humano), por ejemplo, sean el mismo estado²⁸. Este argumento, entonces parte de imaginar una situación en la que un estado físico determinado es el caso, pero en la que el carácter fenoménico que suele acompañarle está ausente, esto es, una situación en la que la estimulación de fibras C, siguiendo con el ejemplo, se da sin la presencia de dolor. Esto es a grandes rasgos lo que “concebir” significa, pero se pueden hacer aún más precisiones acerca de esta noción técnica.

De acuerdo con Chalmers (2002a), quien es uno de los defensores más famosos del argumento de la concebibilidad²⁹, para saber cuál es el sentido relevante de “concebibilidad”, podemos hacer una clasificación en tres dimensiones de las distintas nociones usuales. Primero, la concebibilidad puede ser *prima facie* o ideal. Algo es concebible *prima facie* cuando es imaginado como posible en un primer momento, pero quizás una reflexión más profunda muestre lo contrario. Por otro lado, algo es idealmente concebible cuando incluso un sujeto ideal, sin limitaciones cognitivas, concibe algo como posible, en este caso ninguna reflexión futura retractaría la concebibilidad. Es la concebibilidad ideal la que es pertinente en este argumento.

²⁸ Al igual que los ejemplos acerca del agua, los ejemplos acerca del dolor y las fibras C son un clásico en la literatura y he decidido conservarlos.

²⁹ Otra famosa defensa del argumento de la concebibilidad la realiza Kripke en la tercera conferencia de *El nombrar y la necesidad* ((1980) 2017). En esta formulación, el argumento pretende dirigirse al fisicalismo en tanto tesis ontológica. La versión de Chalmers, en tanto inscrita en un marco bidimensionalista, permite separar con claridad dos pasos en el argumento, siendo el primero de ellos el que es en contra de las explicaciones reductivas y el segundo en contra del fisicalismo. Es por ello que elegí exponer esta versión.

En la segunda dimensión de la clasificación, la concebibilidad puede ser positiva o negativa. algo es positivamente concebible cuando es posible imaginar una situación en la que eso sea el caso, en dónde imaginación debe entenderse en un sentido más amplio que formarse una imagen visual mental (por ejemplo, es concebible que no exista la FFyL de la UNAM, sin que esto corresponda a una- o una única- imagen mental). En contraste, algo resulta negativamente concebible cuando su posibilidad no puede rechazarse. La noción de concebibilidad negativa puede tener distintos matices, pero la versión filosóficamente más pertinente, en particular para este argumento, es entender como negativamente concebible algo cuya posibilidad no puede rechazarse desde lo que sabemos *a priori*; en contraste a aquello que no puede rechazarse partiendo de algún tipo de conocimiento empírico. La concebibilidad positiva implica la concebibilidad negativa, pues si algo puede imaginarse, entonces no puede rechazarse *a priori*, siendo así la primera más fuerte que la segunda, aunque cualquiera de las dos versiones es útil para el argumento.³⁰

Finalmente, la concebibilidad puede ser primaria o secundaria. Esta es la dimensión más compleja de la clasificación, y se basa en la distinción entre intensiones primarias y secundarias de los términos involucrados, expuesta en el capítulo anterior. Estos dos tipos de concebibilidad originan también, de acuerdo con Chalmers, dos tipos distintos de posibilidad y necesidad.

Recapitulando lo dicho en el capítulo segundo, de acuerdo con la intensión primaria de los términos “agua” y “H₂O” es concebible una situación en la que sus referentes se den uno sin el otro, por ejemplo en la tierra gemela en dónde agua es XYZ. En cambio, siguiendo sus intensiones secundarias, no es concebible una situación en la que esto ocurra, pues, como se explicó, agua es necesariamente H₂O. El primer caso es la concebibilidad primaria, mientras que el segundo es la

³⁰ En Chalmers (2002a) hay aclaraciones detalladas sobre distintas nociones problemáticas en esta clasificación, tales como “idealización cognitiva”, “imaginación modal” (el tipo de imaginación que considera pertinente en la concebibilidad positiva) y “a prioridad”. Aunque Chalmers reconoce lo problemático de estos conceptos, apela muchas veces a intuiciones y aproximaciones; lo mismo hago aquí pues discutir cada una queda fuera de mis propósitos.

concebibilidad secundaria. A su vez, la posibilidad primaria, surgida dentro de la concebibilidad primaria, es una posibilidad epistémica; mientras que lo que se concibe siguiendo las intensiones secundarias empata con las posibilidades metafísicas. Es la concebibilidad primaria, en tanto su dimensión epistémica y *a prioridad*, la que interesa para los debates en torno a la explicación.

Así, el argumento de la concebibilidad busca sostener que son primaria e idealmente concebible, ya sea positiva o negativamente, situaciones en las que los sucesos, estados o propiedades físicas se presentan sin los sucesos fenoménicos que comúnmente asociamos o identificamos con los mismos. Específicamente, busca establecer como concebible una situación en la que P , el conjunto completo de verdades físicas de nuestro mundo, sea verdadero y Q , una verdad fenoménica arbitraria que busca explicarse, no lo sea. Dado el carácter *a priori* de esta concebibilidad, $P \rightarrow Q$ no sería *a priori* y la tesis de la implicación *a priori* (TIA), quedaría falseada.

Hay que hacer una aclaración importante. Supongamos que el dolor es idéntico a la estimulación de las fibras C, esta identidad es *a posteriori*, igual que la identidad entre agua y H_2O , o entre el calor y el movimiento cinético molecular, etc. Sin embargo, se ha dicho que es posible imaginar que estas identidades no sean el caso, que agua sea XYZ, por ejemplo. ¿Qué hace diferente esta concebibilidad respecto de concebir que el dolor no sea la estimulación de fibras C? Pues bien, lo que se juega en el argumento de la concebibilidad es el imaginar situaciones a partir de un escenario determinado específico, descrito por P , en este caso. Desde P , un mundo que es indiscernible al nuestro en su distribución física, no es concebible que el agua no sea H_2O , pero sí que el dolor no sea la estimulación de fibras C.

Visto de otra forma, mientras que es concebible que Agua no sea H_2O , lo inverso no es el caso, no es concebible que H_2O no sea agua; en el caso de los estados fenoménicamente conscientes, en contraste, la concebibilidad se mantiene en ambas direcciones. Es con este énfasis

que el argumento de la concebibilidad funciona y hace del caso de la conciencia diferente a otras identidades teóricas.

La forma de concretizar esto ha sido mediante el experimento mental de los llamados Zombies fenoménicos. Un zombie fenoménico es muy diferente a un antropófago muerto viviente de la cultura pop; de hecho, el zombie fenoménico es físicamente indiscernible a los humanos, en cada partícula básica, y, en tanto las propiedades físicas sean causantes de la conducta, su comportamiento será también similar. La única diferencia entre esta criatura y los humanos radica en que el zombie no tiene experiencias conscientes, es decir, ninguno de sus estados mentales posee un carácter fenoménico. Así, la figura del zombie concretiza la idea de que es concebible una situación en la que lo físico y fenoménico sean conceptualmente distintos y disociables.

Este argumento puede formularse con distintos niveles de fuerza. El más fuerte consiste en concebir, en el sentido mencionado como relevante, un mundo que es un duplicado microfísico del actual, pero en el que todos (o gran parte) de los hechos fenoménicos de nuestro mundo están ausentes. Así, los duplicados zombies humanos, animales o de cualquier organismo que en el mundo actual tenga experiencias fenoménicas de algún tipo, carecerían de todas (o gran parte de) ellas.

Algunas versiones menos exigentes, pero igualmente efectivas, consisten en concebir no un mundo entero, sino un duplicado zombie de un individuo (o más bien, un mundo en el que todo es indiscernible excepto este individuo). Así, podríamos concebir un duplicado zombie de mí mismo, indiscernible a nivel físico, pero sin sensación fenoménica alguna. O bien, en versiones aún más débiles, podría sostenerse que mi duplicado microfísico no sea más que un zombie parcial, es decir, que aunque goza de algunas experiencias fenoménicas, carece de algunas otras; o incluso un duplicado de espectro invertido, es decir, un individuo cuyas experiencias fenoménicas sean diametralmente opuestas a las mías (digamos que tiene la sensación que yo tengo al tener orgasmos mientras está en el estado físico en el que yo me encuentro al experimentar dolores, por ejemplo),

en donde el estado fenoménico ausente o invertido es aquel que busca explicarse. Todos estos escenarios apoyan la concebibilidad de mundos en los que $P \wedge \neg Q$, para alguna Q verdadera en el mundo actual que buscamos explicar y que concebimos como ausente, como parte del conjunto de ausentes, o como invertida.³¹

Debe insistirse en que el argumento, en tanto ataca la a prioridad de la TIA, requiere solo que el Zombie sea concebible, no necesita comprometerse con que estas criaturas pudieran existir en el mundo actual, o incluso con que sean metafísicamente posibles³², lo cual podría ser falso. Así, el énfasis está puesto en la *a prioridad*, no en si, dada cierta información empírica acerca de la relación entre estados físicos y fenoménicos, resulta naturalmente imposible que exista un zombie.

Desde mi punto de vista, y concuerdo con Chalmers y los defensores de la concebibilidad, el zombie resulta concebible y congruente. Además desde las versiones menos exigentes, el zombie parcial y el de espectro invertido, se puede avanzar gradualmente a las versiones más fuertes sin que sea claro que en algún punto exista incongruencia alguna. Rechazar que el zombie fenoménico sea concebible requeriría un análisis conceptual (no empírico) de los estados fenoménicos que muestre que estos son también estados físicos sin que este sea *ad hoc*.

Con todo esto, la carga de la prueba recae en quienes rechacen el argumento de la concebibilidad. De cualquier forma, si el Zombie resulta problemático, existe un argumento más en contra de la tesis de la implicación *a priori* para estados fenoménicos: el argumento del conocimiento.

3.2 El argumento del conocimiento: Murciélagos y neurocientíficas presas.

³¹ Así, algunas versiones del argumento de la concebibilidad pueden entenderse como una versión específica del problema de las otras mentes.

³² Argumentar en contra del fisicalismo, en contraste, requiere efectivamente que el zombie sea metafísicamente posible, pero para argumentar contra las explicaciones reductivas basta con que sea concebible (véase Chalmers, 2002a y 2009 para un argumento de la concebibilidad a la posibilidad metafísica).

La formulación más famosa del llamado argumento del conocimiento se la debemos a Frank Jackson (1982, 1986), aunque otra conocida formulación se encuentra en Nagel (1974). Este argumento, igual que el de la concebibilidad, ataca la tesis de la implicación *a priori*, pero lo hace apelando a intuiciones epistémicas, pues sostiene que es posible conocer *P* sin deducir de allí *Q*, con lo que $P \rightarrow Q$ no sería *a priori*.

Para mostrar esto tanto Jackson como Nagel se valen de experimentos mentales. En su famoso artículo *What it is like to be a bat?* (1974), después de introducir la noción de conciencia fenoménica, Nagel formula un experimento mental con el que muestra lo problemática que es ésta para las explicaciones reductivas. Éste ejercicio consiste en intentar saber cómo debe sentirse ser un murciélago, específicamente cómo son las sensaciones obtenidas al operar con un sistema de ecolocalización, partiendo de toda la información física que podamos tener acerca de estos sistemas perceptuales y de la constitución cerebral y del sistema nervioso de los murciélagos. Simplemente no tenemos forma de saber *cómo es* para el murciélago ser un murciélago. Esto es, aunque conozcamos la parte de *P* relevante para el caso, no podemos deducir *Q* (en donde *Q* es una verdad acerca de la fenomenología perceptual ecolocalizadora del murciélago).³³

Aunque el experimento de Nagel es *prima facie* muy convincente, en este trabajo me centraré en el experimento mental de Jackson (1982) ya que, además de ser la formulación que más se ha debatido (incluso Jackson (1986) ha contestado algunas críticas), considero tiene más poder explicativo y alcance argumentativo, pues se enfoca en la percepción humana.

Jackson propone imaginar una situación en la que una neurocientífica, Mary, especializada en fisiología de la percepción visual, ha pasado toda su vida encerrada en una habitación en donde

³³ LA elección del murciélago no es arbitraria. De acuerdo con Nagel, el murciélago es suficientemente cercano al humano (ambos son mamíferos) como para negarle atribución de propiedades fenoménicas; pero es, al mismo tiempo, suficientemente lejano para tener experiencias de un tipo totalmente distinto.

todo se encuentra en escalas de grises. Incluso, de alguna forma, al verse las extremidades o en el espejo no ve tampoco otra gama de colores.

No obstante, Mary conoce todo lo físico que hay que saber respecto a la percepción visual. En otras palabras, posee todos los conocimientos de una hipotética ciencia cerebral acabada, pero también sabe acerca de las amplitudes de ondas y transmisión de la luz, etc. Todo este cúmulo de conocimientos fue obtenido mediante lecciones en televisores, libros, y otros recursos didácticos también en escala de grises. Mary ha tenido también oportunidad de experimentar y manipular cerebros, quizá incluso sobre algunos vivos, pero a estos, siguiendo la hipótesis del experimento, también los vio grises.

Sin embargo, en algún momento Mary logra salir de su acromatópsica reclusión. Al encontrarse en el exterior, tendrá finalmente experiencias de color. La intuición de Jackson es que Mary aprende algo al salir de su reclusión, adquiere nueva información. Si esta intuición resulta ser correcta, como *ex hypothesi* Mary poseía toda la información física relevante, entonces lo que aprende no es información física. El argumento, en palabras de Jackson:

- 1) Mary (antes de su liberación) conoce todo lo físico que hay por conocer acerca de otras personas.
 - 2) Mary (antes de su liberación) no conoce todo lo que hay que conocer acerca de otras personas (Porque ella *aprende* algo acerca de ellos al ser liberada).
- Por lo tanto,
- 3) Existen verdades acerca de otros (y acerca de ella) que escapan a la descripción fisicalista. (Jackson, 1986; 293. Traducción mía, énfasis en el original)

En otras palabras, Mary conocía *P* antes de su liberación, pero no conocía *Q*, siendo *Q* alguna descripción acerca de la fenomenología visual suya o de otros, ni podía deducirlo *a priori* desde su gris sillón. Es hasta su liberación que la neurocientífica aprende *Q*.

Aunque la intuición detrás de la formulación del argumento del conocimiento de Jackson es también muy fuerte, esta ha sido sujeto de distintas críticas. Jackson (1986) mismo ha hecho aclaraciones respondiendo a algunas, en particular las formuladas por Churchland (1985). Antes de

exponerlas, hay que mencionar que el argumento de Jackson se dirigía originalmente al fisicalismo, por lo que algunas de sus críticas se dirigen a este aspecto³⁴. Evaluar estas críticas escapa a los límites de este trabajo.

La crítica más relevante que consideraré parte de una distinción entre dos tipos de “conocimiento”. De acuerdo con esto, si bien Mary aprende algo al salir de su reclusión, esto no sería un conocimiento en el mismo sentido en que lo era el que poseía durante su encierro (conocimiento proposicional), es decir, no adquiere nueva información referente a otros estados, sino algo cercano a una habilidad de reconocer un mismo estado bajo otra presentación (Lewis, 1990), o un “saber cómo” y no un “saber qué” (Churchland, 1985).

Sin embargo, Jackson (1986) contesta que esto efectivamente podría ser el caso si aquello que Mary adquiriera fuese exclusivamente conocimiento acerca de su propia experiencia fenoménica, pero Mary aprende también acerca de la riqueza de la experiencia visual de otras personas, como puede notarse en la formulación de su argumento, algo que desconocía. Este nuevo conocimiento no es una habilidad, sino que está más cercano al conocimiento proposicional que poseía antes de salir de su reclusión.

Pero más aún, este tipo de crítica es también formulada respecto a las consecuencias ontológicas del argumento, pero podría no afectar sus consecuencias explicativas. A grandes rasgos, lo que se intenta es argumentar que el “conocimiento” adquirido por Mary, a pesar de poder

³⁴ Famosamente la crítica conocida como “paridad de razones” sostiene que si el argumento es efectivo en contra del fisicalismo, lo sería, por las mismas razones, en contra del dualismo. Esta crítica propone invertir la situación del experimento mental, tal que Mary conoce todo lo no-físico (suponiendo el dualismo) que hay que saber acerca de la fenomenología visual. De acuerdo con los proponentes de la paridad de razones, la intuición acerca de que Mary aprende algo nuevo se conserva en este caso, resultando en un argumento efectivo también en contra del dualismo.

Jackson contesta que, contrario al conocimiento físico, no es claro que el conocimiento no-físico pueda adquirirse en una situación de reclusión como la del argumento original. Sin embargo, aunque la paridad de razones sea el caso, resulta irrelevante para las conclusiones respecto a las explicaciones reductivas. La consecuencia relevante, a lo más, sería sólo que no son posibles las explicaciones reductivas de la conciencia fenoménica a partir de lo que sea que resulte estar a la base de ella en alguna versión del dualismo (digamos, alguna propiedad protofenoménica. Véase nota 9).

llamarle “información fenoménica”, no tendría que ser necesariamente conocimiento referente a propiedades no-físicas, sino sólo un tipo de conocimiento que las presenta de distintas maneras. Aún en esta circunstancia, podría ser suficiente para rescatar la tesis de la implicación *a priori* que esta forma de presentación no es deducible *a priori* de *P*.

Así, podría argumentarse también, y en relación a lo anterior, que lo que hace falta a Mary es adquirir los conceptos relevantes (aquellos en los que se presenta *Q*), para poder satisfacer la TIA, los que, sin embargo, solo adquiere al tener la experiencia. Esto es un tema sobre el que se ha debatido mucho recientemente y en el que no parece que el consenso se logre próximamente. Profundizar en el debate exigiría más capacidades conceptuales de las que tengo en este momento, con lo cual en esta sección me basta con apoyar la plausibilidad –aunque no una conclusión definitiva- del argumento del conocimiento.

Tal como vimos en el capítulo primero, la satisfacción de la tesis de la implicación *a priori* requiere de un análisis conceptual. Sin embargo, en la siguiente sección, a través de una distinción entre conceptos psicológicos y fenoménicos, explicaré que los conceptos fenoménicos no pueden ser analizados funcionalmente de manera relevante.

Antes de continuar, vale la pena señalar que otros fenómenos naturales parecen ser efectivamente implicados *a priori* por *P*. Concebir una situación en la que *P* se mantiene y hechos como la identidad agua y H₂O, o que el agua cubra el 60% del planeta, parecen, al menos, más difíciles de concebir que los escenarios zombies, como ya se mencionó, y que en posesión del concepto “agua” son deducibles de *P*.

3.3 Conceptos fenoménicos y psicológicos: Dos formas de referir a la mente.

Siguiendo a Chalmers (1996; 3-32) existen sólo dos tipos de conceptos mediante los cuales referimos a nuestros estados mentales: conceptos fenoménicos y conceptos psicológicos. Los primeros, evidentemente, los usamos para referir al aspecto sensitivo, al cómo es estar en ellos; los segundos, como su nombre sugiere, para referir a nuestros estados mentales en tanto éstos juegan un papel en la explicación del comportamiento o conducta.

De acuerdo con el funcionalismo analítico (por ejemplo, Lewis (1996)), todos los conceptos mentales serían psicológicos, es decir, términos funcionales definidos por sus roles causales (no debe confundirse con el uso teleológico de “funcional”). Sin embargo, los estados fenoménicos escapan a este tipo de análisis, pues, a grandes rasgos, explicar el papel que un estado desempeña en la conducta no dice nada acerca de porque posee determinado carácter fenoménico y no otro, o ninguno en absoluto. En esta sección expondré y apoyaré esta distinción.

De acuerdo con Chalmers, el análisis de nuestros estados mentales se agota en estos dos tipos de conceptos. Por ejemplo las creencias, deseos, u otros estados intencionales pueden describirse mayoritariamente usando conceptos psicológicos. Lo mismo ocurre con términos que refieren a procesos mentales como “aprendizaje” o “memoria”, el primero es aproximadamente el proceso por el que se adquiere nueva información, mientras que el segundo define a algún proceso de almacenamiento de información adquirida y su disposición para posterior uso (Evidentemente estos análisis son muy básicos, pero pueden construirse análisis más sofisticados y, en principio, completos). Y, más aún, los términos de clases naturales como “agua”, “calor”, etc. tienen también análisis funcionales.

Muchos de nuestros términos mentales ordinarios tienen un doble análisis conceptual. Por ejemplo, “dolor” puede analizarse funcionalmente como el tipo de estado que es causado por algún tipo de irritación sensorial (por ejemplo) y típicamente causa algunas conductas como gritos, quejas, movimientos, etc. dependiendo en gran medida de sus interacciones con otros estados mentales.

Sin embargo, “dolor” también puede referir a la incómoda sensación que suele acompañar esos roles causales. Cotidianamente por “dolor” nos referimos a una mezcla de estos dos significados, pero podemos hacer una distinción entre “dolor_f” (fenoménico) y “dolor_p” (psicológico).

Si bien muchos términos mentales son como el anterior y tienen, en mayor o menor medida, un componente psicológico y uno fenoménico, es filosóficamente pertinente sostener una distinción conceptual entre estos componentes, es decir, podemos imaginar que exista algún estado encargado de satisfacer el rol causal por el que definimos el concepto psicológico, pero en el que esté ausente la sensación a la que refiere el concepto fenoménico que lo acompaña o viceversa. Continuando con el ejemplo, podemos imaginar un estado físico que satisface el rol causal de “dolor_p”, es decir, un estado causado por alguna irritación que causa tales o cuales comportamientos, y que sin embargo carezca de cualquier carácter fenoménico. Este estado puede ser el referente del concepto “dolor_p”, pero no del fenoménico, pues el componente sensorial está ausente.

Precisamente por esta independencia conceptual, ningún concepto nos dice nada acerca del otro. En particular, las definiciones funcionales de los estados psicológicos no dicen nada acerca de porque estos estados se sienten de determinada manera y no de otra, o bien porque se sienten de alguna manera y no de ninguna en absoluto. Es decir, los términos psicológicos y fenoménicos no son ni interdefinibles ni intercambiables.

De esta forma, existen también dos tipos de *explanandum* de los que dar cuenta. Podemos pedir una explicación de un aspecto mental psicológico o de uno fenoménico; explicar el primero no es explicar el segundo, ni viceversa. Es, reitero, del aspecto fenoménico del que me interesa evaluar la posibilidad de explicaciones reductivas.

Como se concluyó en el capítulo segundo, una explicación reductiva funciona por el análisis conceptual de los términos en los que se describe el *explanandum*. Como también se vio, el análisis

de estos términos es del tipo funcional, a partir de las intensiones primarias. Sin embargo, de un concepto fenoménico no se puede hacer ningún análisis pertinente sin que este sea circular (p. ej. “sensación de rojo_f” puede analizarse como “el estado que se siente de tal y cual forma (cómo ver rojo, de hecho)”), o bien como acompañante tradicional de un estado psicológico o físico (“El estado fenoménico que suele presentarse cuando el organismo está en un estado psicológico/físico en el que tal y cual ...”), pero esto no dice nada acerca de porque el carácter fenoménico suele acompañar la función psicológica o al estado físico.

Ninguna de estas opciones permite una explicación reductiva, pues lo que podemos mostrar a través de la evidencia empírica es cómo ciertos procesos microfísicos se encargan de satisfacer los roles causales de los conceptos psicológicos. Pero en el caso de los conceptos fenoménicos, no hay un rol claro qué satisfacer.³⁵ Por ejemplo, una explicación reductiva acorde al modelo ND de dolor_p sería la siguiente:

P1. S está en un estado cerebral de estimulación de fibras C

P2. “dolor_p” =_{def} El estado causado comúnmente por tal y cual ... y que suele causar...

P3. La estimulación de fibras C cumple con ser el referente del análisis de “dolor_p” (al menos en el caso de los sistemas del tipo que S es)

C. S está en un estado de dolor_p

Mostrar que un estado físico cumple cierto rol funcional, tarea empírica para las ciencias empíricas, expresada principalmente en la premisa 3, no es equivalente a mostrar que este estado es el encargado de cumplir con algún carácter fenoménico. Dados los argumentos del conocimiento

³⁵ La literatura en torno a los conceptos fenoménicos es bastante amplia. Recientemente se ha puesto el énfasis en ellos pues se ha debatido pueden explicar la brecha explicativa de forma compatible con el fisicalismo (esto es la llamada estrategia de los conceptos fenoménicos, bautizada así por Stoljar (2005), pero cuyos antecedentes pueden rastrearse hasta Loar (1990)). Para un mapa de las distintas propuestas de cómo son los conceptos fenoménicos, puede verse Balog (2009); para mis propósitos es suficiente sostener que los conceptos fenoménicos no poseen análisis funcionales significativos.

y la concebibilidad y la distinción conceptual expuesta, después de explicar porque S se encuentra en un estado de dolor_p, sigue siendo válido preguntar si S se encuentra en un estado de dolor fenoménico.

Así, no podría darse una explicación del dolor fenoménico de este tipo, pues los conceptos fenoménicos no pueden analizarse de la forma en que la premisa dos requeriría.

En este capítulo he mostrado los argumentos que nos llevan a concluir que la presencia de la conciencia fenoménica no puede explicarse apelando puramente a estados físicos. Si bien no siempre es clara la dialéctica entre estos argumentos, es decir, si es a causa de la distinción conceptual que nuestras intuiciones respecto al argumento del conocimiento y de la concebibilidad son tales, o bien estas intuiciones respecto a estos argumentos apoyan la distinción conceptual; el hecho de que estos argumentos se refuercen mutuamente, hace difícil que una sola crítica sea totalmente socavadora. Aunque no he mostrado, ni he pretendido mostrar, que estos argumentos sean concluyentes, pues solo he expuesto pocas críticas (y por supuesto, otras tantas pueden seguir esgrimiéndose) considero que dan buenas razones para considerar que la conciencia no puede ser explicada reductivamente.

En el siguiente capítulo, expondré algunas alternativas propuestas como explicaciones reductivas, aún dentro del modelo ND, desde las cuales, se argumenta, la conciencia podría explicarse reductivamente, aun con la falta de análisis conceptual de sus términos. Aunque la conciencia puede explicarse conforme a estos modelos, si tengo razón en lo expuesto, estos modelos no pueden considerarse reductivos pues incumplen con (R) y con la tesis de la implicación *a priori*.

Capítulo 4:

Las explicaciones no reductivas de la conciencia fenoménica.

En el capítulo anterior expuse los argumentos que nos inclinan a pensar que la conciencia no puede ser sujeto del modelo de explicación reductiva propuesto. Esto en gran medida por la falta de análisis conceptual de los conceptos fenoménicos, que impide se cumpla la tesis de la implicación *a priori* y con ello (R).

En este capítulo expondré dos propuestas alternativas de explicación reductiva, aún dentro del modelo ND, que se argumenta no necesitan el análisis conceptual expuesto en el capítulo 2: las leyes puente del modelo nageliano, en la sección 4.2, y las identidades *a posteriori*, en la sección 4.3. Si estas alternativas, que comparten mucho en cuanto a su estructura, resultasen satisfactorias, el caso de la conciencia no sería especialmente problemático, pues aún sin ofrecerse el análisis conceptual habría alternativas para formular una explicación reductiva. Estas propuestas ponen en cuestión la necesidad del análisis conceptual para las explicaciones. Pero antes de presentarlas, en la siguiente sección expondré una crítica más directa hacia el papel del análisis conceptual en las explicaciones reductivas hecha por Block y Stalnaker.

Aunque la presencia de la conciencia en estados particulares se puede explicar siguiendo estas alternativas, suscribiré una crítica a estas posturas que las considera como explicaciones no reductivas y las sitúa en desventaja explicativa respecto al modelo basado en el análisis conceptual.

Con lo cual expondré también como es el tipo de explicación disponible para la presencia de la conciencia, siendo estas explicaciones no reductivas, de allí el título del capítulo.

4.1 Block y Stalnaker en contra del análisis conceptual

En el capítulo segundo expuse la tesis de la implicación *a priori*, su relación con el modelo ND de explicación científica reductiva y cómo, de acuerdo con Chalmers y Jackson (2001), esta es satisfecha a partir de un tipo determinado de análisis conceptual. Ahora bien, esta tesis ha sido criticada célebremente por Block y Stalnaker (1999), quienes argumentan deflacionariamente. Según ellos aun si los argumentos expresados en el capítulo anterior resultaran ciertos, la conciencia no sería un caso especialmente problemático, pues si bien es cierto que de los conceptos fenoménicos no puede ofrecerse análisis conceptual, tampoco esto sería el caso para la mayoría de los términos de clase natural.

Aunado a esto, señalan que en de hecho las brechas explicativas³⁶ se cierran de forma fundamentalmente empírica, y formulan una propuesta de explicación reductiva basada en cómo de hecho, de acuerdo con ellos, se opera en las ciencias. De esta forma, puede dividirse el programa crítico de Block y Stalnaker en dos pasos principales, aunque íntimamente relacionados. El primero es criticar la noción de análisis conceptual; el segundo, dar un alternativa.

En esta sección expondré brevemente el núcleo de sus críticas y como Chalmers y Jackson (2001) las contestan. En las secciones siguientes expondré su propuesta, relacionándola también con el modelo nomológico deductivo, y las críticas a las que se enfrenta. Así, concluiré que su postura es difícilmente sostenible y que la conciencia es un caso especialmente problemático.

³⁶ Block y Stalnaker utilizan mayoritariamente el término “brecha explicativa”, pero en realidad con lo que discuten es con las explicaciones reductivas tal como las he formulado aquí.

Para Block y Stalnaker, lo que permite cerrar una brecha explicativa no es necesariamente un análisis conceptual. Lo que se deduce en un argumento deductivo es la descripción de un fenómeno en términos microfísicos, lo que se necesita para establecer una conexión con los términos de distinto nivel en los que se presentó el fenómeno a explicar es simplemente establecer una identidad, en sus palabras:

Lo que se deduce de hecho en tal explicación es una descripción de aquello a explicar enteramente en el lenguaje de la ciencia. Para responder la cuestión explicativa original, presentada en vocabulario ordinario, todo lo que tenemos que agregar es la premisa de que el fenómeno descrito en lenguaje científico es el mismo que el presentado en lenguaje ordinario, pero descrito de otra forma” (Block y Stalnaker, 1999; 9. Traducción mía).

Block y Stalnaker sostienen que si bien el análisis conceptual podría ser suficiente para proveer una explicación reductiva, éste no es necesario. La muestra de ello es que de la mayoría (y quizá de todos) los términos que refieren a clases naturales no podría darse un análisis conceptual apropiado y, sin embargo, la ciencia ha provisto explicaciones reductivas satisfactorias.

De acuerdo con estos autores, un análisis conceptual, por ejemplo del término “calor” necesita evitar los casos del calor fantasma (*ghostheat*), es decir aquellos en los que hay una substancia no-física que satisface también el rol conceptual asociado al calor, haciendo que la identidad entre calor y movimiento cinético molecular sea falsa (lo mismo en casos de agua fantasma o *ghostwater*). Igualmente, el análisis necesita garantizar la unicidad del referente. La forma en que un análisis conceptual podría garantizar esto es que este análisis sea algo muy cercano a una definición en términos de suficiencia y necesidad, tal que esté libre de contraejemplos (incluso de aquellos que sean sólo epistémicamente posibles, como el calor fantasma que claramente no existe en el mundo actual).

No obstante, no es posible formular análisis conceptuales tan fuertes. De acuerdo con Block y Stalnaker, los análisis conceptuales ofrecidos por Chalmers y Jackson no establecen condiciones

de necesidad ni suficiencia. Por ejemplo, para el significado de “vida”, un ejemplo ofrecido por Chalmers (1996; 79), se ofrece algo basado en las funciones de respiración, nutrición, excreción, movimiento, reproducción, etc. Pero es fácil imaginar organismos vivos que no satisfacen alguna de ellas, o bien organismos no vivos que cumplan con algunas. Con esto, de acuerdo con Block y Stalnaker, no se podría garantizar la identidad entre los referentes.

Sin embargo, de acuerdo Chalmers y Jackson, estas críticas están basadas en una concepción de análisis conceptual mucho más fuerte de lo que se requiere. Tal como mencioné en el capítulo segundo, el análisis conceptual relevante es un análisis funcional-causal aproximado, que muchas veces no es explícito, cercano a la noción de intensión primaria y que no tiene por qué estar libre de corrección y refinamiento conceptual al considerar mayor evidencia empírica que queramos incluir bajo la extensión del concepto.

Por otro lado, la unicidad del referente de los términos, cuando esto es el caso, es garantizada por la cláusula (T) agregada al antecedente del condicional (aunque es quizá cierto que no por el análisis mismo); la misma cláusula evita que existan referentes no-físicos de algún término, (es decir, evita casos como el del calor o agua fantasma). Pero, como ya se dijo al final del capítulo segundo (págs. 40-45) esto también depende de los términos que se tomen como entrada de la función, si se considera un mundo en el que hay H₂O y XYZ, quizá no sea necesario garantizar la unicidad del término “agua” para ese escenario. Si el análisis conceptual tuviese que ser de la forma en la que Block y Stalnaker lo entienden, quizá tendrían razón al señalar que proveer explicaciones reductivas sería una tarea casi imposible.

En cuanto a la propuesta positiva de Block y Stalnaker, éstos sostienen que la explicación científica reductiva se fundamenta empíricamente, lo que permite justificar las identidades son criterios de metodología científica. Es esta propuesta la que me interesa presentar y contestar en este capítulo, pero antes de hacerlo dedicaré una sección al paradigmático modelo de explicación

reductiva de Ernest Nagel, ya que la propuesta Block y Stalnaker puede entenderse mejor como un caso específico de lo que dice este modelo, pues ambos comparten en gran medida la misma estructura.

4.2 El modelo nageliano de explicación reductiva.

Ernest Nagel es quizá quien inicia las reflexiones filosóficas en torno a las explicaciones reductivas, o por lo menos sus consideraciones en este tema han sido paradigmáticas. Nagel (1961, Cap. XI) concibe las reducciones como una relación entre teorías o leyes científicas³⁷, tal que una teoría, o las leyes que la componen, se reduce a otra si puede deducirse lógicamente de esta. A menudo la deducción no es directa pues los términos involucrados en las teorías no son los mismos, por lo que se hacen necesarias premisas extras que actúen como traducciones entre ellos. Este tipo de reducciones es lo que Nagel denomina reducciones no-homogéneas o heterogéneas, en oposición a las reducciones homogéneas en las que se puede prescindir de estos enunciados traductores.³⁸

Estas proposiciones extra, encargadas de correlacionar los términos entre teorías, reciben el nombre de leyes puente (*bridge laws*) y son el núcleo del modelo Nageliano. Nagel concibe, generalmente, estas leyes puente como bicondicionales, en donde en cada uno de los lados se encuentra un término primitivo de las teorías reductora, de un lado, y por reducir, del otro³⁹. De tal forma estas leyes puente son un tipo de “definiciones” coordinadoras (como suele denominárseles

³⁷ Nagel (1961;) ejemplifica esto con la reducción de la termodinámica a la mecánica estadística a través de mostrar que la ley de Charles-Boyle es una consecuencia de las leyes de la mecánica.

³⁸ Estas deducciones homogéneas no son problemáticas, además quizá no deban ser consideradas como reducciones si tanto el *explanans* como el *explanandum* están expresados en el mismo vocabulario. Por ello, me centraré en exponer el caso de las reducciones no homogéneas, en las que Nagel mismo se centra.

³⁹ Aunque no me es claro que Nagel haya definido siempre las leyes o principios puente como bicondicionales, esta es la forma tradicional en que se las interpreta y discute.

también). Así, estas leyes puente permiten que la deducción opere, o, en caso contrario, permiten reescribir las leyes/teorías que buscaban reducirse y así poder agregarlas a la teoría reductora.

Aunque el reduccionismo Nageliano fue concebido originalmente como una relación entre teorías y leyes científicas, el énfasis que pone en la deducción permite extender su modelo para tratar con las explicaciones reductivas de forma más cercana a lo que aquí me interesa, es decir, a explicaciones de sucesos particulares. Así, podemos definir una explicación reductiva de acuerdo con el modelo Nageliano como aquella en la que, aún dentro del modelo ND, se incluye bicondicionales que operan como leyes puente/ principios coordinadores y que juegan un papel determinante en la deducción del *explanandum* (aunque esto último no sea tanto una teoría como un suceso particular). De tal forma, una explicación de la presencia de un estado fenoménico como dolor, siguiendo el modelo nageliano tendría la siguiente forma:

1. S está en un estado de activación de las fibras C
2. Activación de las fibras C ↔ dolor

Por lo tanto, S está en un estado de dolor

Es la premisa dos la que incorpora una ley puente en la que se correlacionan un estado físico y uno fenoménico y que podría estar justificada directamente por la evidencia empírica obtenida en neurociencias, por ejemplo.

La deducción anterior es válida. No obstante, las leyes puente han sido objeto de distintas críticas, aquí mencionaré sólo las formuladas por Kim (1998, cap. IV; 2005, caps. IV-V). La primera dificultad a la que se enfrenta el modelo de Nagel es que las leyes puente entran en conflicto con la tesis de la realización múltiple (Putnam (1973), Fodor (1974)), que es altamente plausible. De acuerdo con la tesis de la realización múltiple, algunos de los fenómenos y propiedades de alto nivel, aquellos de los que se encargan las ciencias especiales, pueden ser realizados por distintas propiedades de bajo nivel. En el caso de estados fenoménicos, por ejemplo, un estado como el dolor

puede ser identificado o relacionado con distintos estados cerebrales, dependiendo de la especie, el género o las edades. E, incluso, si pensamos en escenarios posibles (al menos epistémicamente posibles) en los que *aliens* o computadores son capaces de poseer estados de dolor, en estos organismos-sistemas el correlato del dolor, evidentemente, no sería el que es en los humanos.

Si la realización múltiple de estados fenoménicos es el caso, entonces no sería posible formular bicondicionales del tipo “dolor \leftrightarrow Estimulación de fibras C”, pues cualquier estado de dolor relacionado con otro estado físico haría falso este bicondicional. Es decir, si la realización múltiple es el caso, las leyes puente no son posibles en sentido estricto.

Hay algunas respuestas relativamente sencillas para escapar a esta conclusión. Primero, puede sostenerse algún tipo de reduccionismo local, en el que las leyes puente son aplicables a alguna especie determinada (aunque qué tan “local” se debería ser es una cuestión problemática)⁴⁰; así, el principio puente se subordina a un condicional, obteniendo una fórmula como $H \rightarrow$ (Estimulación de fibras C \leftrightarrow dolor), en donde H limita la correlación a los humanos, por ejemplo. La segunda alternativa, es convertir uno de los lados del bicondicional en una premisa disyuntiva, por ejemplo Dolor \leftrightarrow (Estimulación de fibras C \vee ... \vee ... \vee ...), aunque esta premisa disyuntiva puede ser inconvenientemente larga. Ambas propuestas son exploradas por Kim (1998; 93)⁴¹

Aunque existan inconvenientes prácticos en estas opciones, en principio permiten que las deducciones funcionen. Sin embargo, vale la pena señalar que las explicaciones reductivas mediante el análisis conceptual no entran en tensión alguna con la tesis de la realización múltiple, pues es perfectamente compatible con que más de un sistema físico cumpla con los roles causales de los

⁴⁰ Por su puesto este es un problema que la información empírica puede solucionar o delimitar. Desconozco si el estado actual de la investigación empírica permite establecer correlaciones físico fenoménicas para el caso de los humanos en tanto especie, o si, por el contrario, la evidencia apoya la conclusión de que habría ser más específicos que eso.

⁴¹ Cabe mencionar que, de acuerdo con Kim, estas premisas disyuntivas no refieren a propiedades disyuntivas, sino que son sólo una disyunción de propiedades, mucho menos problemática.

conceptos en que se presenta el fenómeno a explicar. Esto es ya una ventaja del modelo del análisis conceptual sobre el modelo nageliano.

Pero, más aún, el modelo nageliano enfrenta otras críticas que no son tan fáciles de sortear y que son más relevantes. De acuerdo con Kim, las leyes puente postulan una correlación que no está por sí misma explicada. Esto, en el caso de los estados fenoménicos resulta especialmente problemático, pues, de lo que se trata es de dar cuenta de porqué ocurre un estado fenoménico a partir de ciertos estados físicos, esto es ¿Por qué un estado fenoménico (y no otro, o ninguno) está acompañado, es realizado, o es idéntico a algún estado físico particular (y no otro, o ninguno)? las leyes puente sólo nos recalcan que existe una correlación (o una correlación local) entre estos, sin hacer inteligible ni transparente porque esto así.

A diferencia de las explicaciones mediante leyes puente, el análisis conceptual permite ver cómo es que una propiedad se relaciona con otra a través de la satisfacción de cierto rol funcional que asociamos a los conceptos. Esto último tiene ventajas epistémicas, hace inteligible la relación entre propiedades. En esto, el modelo del análisis conceptual resulta tener ventajas respecto a la transparencia de las explicaciones frente al modelo nageliano.

Finalmente, de acuerdo con Kim, podemos señalar al modelo nageliano como no reductivo, contrario a sus pretensiones. Si bien mediante las leyes puente se puede deducir una descripción de un suceso particular en términos macrofísicos de descripciones en otros términos, para que esto suceda hay que adoptar estos principios coordinadores como básicos en nuestro sistema explicativo. Ya que estos términos son de hecho los mismos usados en el *explanandum*, y las premisas que los expresan son empíricas y no conceptuales, a diferencia de las premisas de análisis conceptual en el modelo favorecido, las leyes puente violan el requisito (R).

Además, aunque quizá menos importante, puesto que las leyes puente se presentan bajo bicondicionales que solo establecen correlaciones, que además descubrimos empíricamente, en

principio éstas son compatibles con ontologías dualistas –por ejemplo un dualismo de propiedades en donde hay correlaciones nomológicas, garantizadas por alguna ley físico-fenomenica, como las que defiende (Chalmers, 1996), o incluso con teorías de la armonía preestablecida como la que suele atribuirse a Leibniz. Aunque por sí mismas no nos dicen que esto sea el caso, las leyes puente podrían no solo agregar nuevos conceptos a nuestras teorías o explicaciones, sino nuevas entidades a nuestra ontología.

Sin embargo, podría argumentarse que estos bicondicionales no reflejan solo correlaciones, sino que de hecho podríamos invocar identidades en una explicación y así evitar engrosar nuestra ontología. Esa es a grandes rasgos la propuesta de Block y Stalnaker (1999) que expondré en la siguiente sección.

4.3 Identidades *a posteriori* y explicaciones reductivas (Block y Stalnaker otra vez).

De acuerdo con Block y Stalnaker, una vez que tenemos las correlaciones (en este caso entre un estado físico y uno fenoménico) estamos justificados por principios de metodología y práctica científica, como parsimonia y simplicidad, en inferir de ella una identidad: la mejor explicación de la presencia de una correlación es una identidad subyacente. Así, una vez obtenida esta identidad, según ellos con criterios empíricos y no conceptuales, podemos usarla como una premisa en la explicación:

1. S está en un estado de activación de las fibras C
2. Activación de las fibras C = dolor

Por lo tanto, S está en un estado de dolor

Lo primero que hay que señalar es que las identidades también están en tensión con la tesis de la realización múltiple, pues, de nuevo, el dolor podría presentarse en organismos que no tienen siquiera fibras C. Así, lo que aquí se toma por identidades debería ser algo menos comprometedor que una identidad entre tipos absoluta; una identidad local (*token*), por ejemplo.

Desde ahora puede observarse que la estructura de esta explicación es muy similar a la que se obtenía al usar leyes puente en la propuesta del modelo nageliano (pág. 64), la diferencia radica en los compromisos y supuestos ontológicos de la premisa 2, mientras que antes era un bicondicional, ahora es una identidad. Así, siguiendo la crítica de Kim, tendrá problemas análogos a los de dicho modelo, en particular respecto a sus alcances explicativos y su relevancia en tanto modelo reductivo.

Existen algunos problemas con el poder explicativo de esta propuesta. De acuerdo con Block y Stalnaker las identidades no explican las correlaciones, sino que actúan deflacionariamente. Es decir, si sabemos que el agua es H₂O se hace innecesario pedir una explicación de por qué cada vez que hay presencia de agua la hay también de H₂O (como innecesario es pedir una explicación de la correlación entre agua y agua). A su vez, las identidades por sí mismas son también inexplicables, no se puede pedir ni ofrecer una explicación de por qué agua es agua, o incluso de por qué agua es H₂O (esto es distinto a explicar cómo de hecho conocemos una identidad, lo que claramente tiene sentido).

Sin embargo, Block y Stalnaker pasan por alto que existe una diferencia entre la mayoría de las identidades teóricas, como "Agua=H₂O", e identidades fenoménicas y estados cerebrales, como "Dolor= Estimulación de fibras C" (asumiendo, claro está, que estas son el caso), y que repercute en los alcances que pudiesen tener al ser usadas como premisas en una explicación. Esta diferencia es señalada ya en un famoso artículo de Levine (1983), pero puede incorporarse y desarrollarse a la luz de lo expuesto por Chalmers y Jackson (2001) en relación a la tesis de la implicación *a priori*.

De acuerdo con Levine (1983), identidades como i) dolor = estimulación de fibras C presentan una brecha explicativa. A diferencia de identidades como ii) Agua = H₂O, i) deja algo inexplicado, una especie de misterio residual. Esta asimetría descansa fundamentalmente en la distinción conceptual que he expuesto en el capítulo anterior.

Siguiendo a Levine, existe una identidad ii') El líquido que causa tal... y que tal = H₂O, en la que del lado izquierdo del signo se presenta el rol funcional que asociamos al término "agua", mientras que del lado derecho el componente químico que descubrimos, empíricamente, ser el único que satisface el rol adecuado en el mundo actual. ii') subyace a ii) y es la que la hace una identidad inteligible pues relaciona H₂O con los mecanismos causales que asociamos al agua. En cambio, "dolor", como se ha señalado no tiene ningún análisis conceptual-funcional pertinente que permita una identidad subyacente análoga a ii'), dejando un aspecto ininteligible en identidades como i). Es ii') lo que permite usar identidades como premisas en una explicación de forma relevante, pues si nuestras identidades fuesen sólo postuladas, no avanzaríamos mucho en la comprensión del *explanandum*.

Esta conclusión puede reformularse y entenderse a partir de lo expuesto en el capítulo segundo respecto a la implicación *a priori*. Al final de dicho capítulo se mencionó que de hecho las identidades teóricas como ii) son implicadas *a priori* por P, nuestra descripción completa del mundo en términos microfísicos, pues dado el análisis conceptual expresado en ii') y la cláusula T, sabemos que en el mundo actual H₂O es la única sustancia que satisface el análisis conceptual. En cambio, identidades en las que se involucran términos fenoménicos, como i), no son implicadas *a priori* por P.

Como ya se ha insistido, tanto i) como ii) son identidades del tipo *a posteriori*, pues tomadas de forma aislada no hay ningún tipo de análisis conceptual de "agua" o "H₂O", o "Dolor" y "estimulación de fibras C", que permita establecer estas identidades; sino que su descubrimiento

es forzosamente empírico. Sin embargo, una vez que tenemos el escenario descrito por *P* y los análisis conceptuales pertinentes, existe una diferencia entre identidades como i) y ii), pues para el caso de las segundas existe una justificación *a priori* subyacente que no está disponible para las primeras (esta justificación es independiente a la forma en que se descubren, que puede ser de hecho empírica). A las identidades que no están implicadas *a priori* por *P*, Chalmers y Jackson (2001; 353-354) las denominan epistémicamente primitivas, pues no son implicadas *a priori* por verdades más básicas.

Esta primitividad epistémica, cabe recalcar, no implica necesariamente una primitividad ontológica, es decir, no implica (al menos no inmediatamente) que los hechos fenoménicos sean hechos no físicos. Es el llamado fisicalismo *a posteriori* (justamente llamado así por aceptar que los hechos fenoménicos no son implicados *a priori* por *P*) la postura que trata de compatibilizar esta conclusión con una ontología fisicalista. Pero como he señalado no entraré en las posibles repercusiones ontológicas de esta conclusión epistémico-explicativa.⁴²

Así, aunque la propuesta de Block y Stalnaker no es totalmente equivocada, pues las identidades *a posteriori* pueden jugar un papel en las explicaciones; este papel es relevante, posibilitando transparencia epistémica, sólo cuando las identidades pueden ser justificadas *a priori* a partir de *P*, lo que no es el caso en las identidades fenoménicas. Así, estas identidades presentan lo que Levine llama brecha explicativa, la cual se proyecta al ser usadas como premisas en una explicación particular, dejando una brecha epistémica, una falta de transparencia, en las explicaciones.

⁴² Chalmers (2003) hace un mapa de diversas posturas disponibles, allí puede encontrarse una exposición del fisicalismo *a posteriori* (que Chalmers llama fisicalismo tipo b). Algunos de quienes sostienen un fisicalismo *a posteriori* son Loar (1990/1997), Hill (1997), Stoljar (2005). Finalmente, un tipo de argumentos de la primitividad epistémica a la ontológica puede verse en Chalmers (2009) y Chalmers (2002a)

Dado que estas identidades no son implicadas *a priori* por *P*, estas tienen que ser incorporadas al antecedente del condicional, esto es, los términos fenoménicos son también incluidos en las premisas explicativas. Ya que esta identidad es *a posteriori*, esto es, no es una definición, ni análisis conceptual, ni enunciado analítico, sino que su verdad se conoce sólo a través de la experiencia, se añade nueva información empírica al *explanans*. Esto es, apelar a las identidades también entra en conflicto con (R) y con la tesis de la implicación *a priori*; pues a diferencia de la ocurrencia del término “dolor” en las premisas de análisis conceptual, en la premisa 2 de la explicación mediante identidades, “dolor” se refiere directamente a una propiedad (igual que en el caso de las correlaciones de las leyes puente).

4.4 Las explicaciones no reductivas de la conciencia fenoménica

Si bien, tanto el modelo de las leyes puente como el de las identidades *a posteriori*, que comparten mucho, pueden de hecho dar explicaciones de la presencia de la conciencia fenoménica, éstas no pueden considerarse modelos de explicaciones reductivas ya que irrumpen con el requisito (R) y con la tesis de la implicación *a priori*. Así, lo que tenemos es un par de modelos de explicaciones no reductivas por las que podemos dar cuenta de la ocurrencia particular de la conciencia fenoménica.

En palabras de Chalmers y Jackson, algún hecho *M* en el que hay presencia de estados fenoménicos es implicado *a priori* por $P \wedge Q$, en donde *Q* suma información fenoménica al antecedente del condicional, pero no por *P* misma. (Chalmers y Jackson, 2001; 352-353). Dependiendo del modelo, esta información adicional vendrá en forma de bicondicional o de identidad (es decir, o bien $Q = (\text{Dolor} \leftrightarrow \text{Estimulación de fibras C})$, o bien, $Q = (\text{Dolor} = \text{Estimulación de fibras C})$).

Podría objetarse, que la argumentación puede ser circular y que bien pueden rechazarse dichos requisitos en pos de aceptar alguna de estas dos propuestas como algún tipo de explicación reductiva. Sin embargo, no me parece que esto sea el caso, pues, primero se ha mencionado que para otros fenómenos naturales los requisitos solicitados pueden ser satisfechos; segundo, y más importante, se ha expuesto que estos requisitos otorgan un tipo de claridad epistémica a las explicaciones, la mínima que podríamos esperar de una explicación reductiva sin trivializar esta noción, y que no es garantizada por ninguna de estas dos alternativas. De hecho, se ha expuesto, las explicaciones no reductivas de la conciencia fenoménica no serían totalmente explicativas, pues no cerrarían la llamada brecha explicativa.

Dicho esto, resultaría inadecuado rechazar un modelo de explicación reductiva que parece útil en muchos casos, aunque no en el de la conciencia, sólo para aceptar otros modelos en los que la conciencia tenga cabida, sobre todo si esto significa desposeer a lo que se considere explicaciones reductivas de ventajas explicativas. Es decir, de esta manera se trivializaría la noción de explicación reductiva.

En conclusión, en este capítulo he expuesto dos modelos explicativos que, contrario a como se proponen, no pueden ser considerados reductivos. Aunque la presencia de la conciencia fenoménica en estados mentales particulares puede explicarse tal como estas alternativas proponen, no puede considerarse que estas sean explicaciones reductivas en términos físicos.

Además de exponer la forma en que tendrían las explicaciones no reductivas de la conciencia, exponer las críticas a estas propuestas alternativas ha servido para reforzar, indirectamente, el modelo ofrecido de explicación reductiva.

Conclusiones

En esta tesis he expuesto el debate en torno a las explicaciones reductivas de la conciencia, inclinándome por un lado del mismo, pero pretendiendo exponerlo de forma general, clara y precisa. Mi hipótesis de partida era que de la conciencia fenoménica no pueden ofrecerse explicaciones de carácter reductivo. Aunque he terminado efectivamente sosteniéndola, al estudiar con detalle el debate me ha resultado menos claro que esta sea una conclusión tajante y última. Como mencioné en la última sección del capítulo, hay aún más argumentos que considerar y, como reconocí en la introducción, muchos de los supuestos y argumentos que presenté no son incontrovertidos.

Para llegar a esta conclusión he presentado un requisito para las explicaciones reductivas, limitando la discusión al modelo nomológico deductivo, lo relacioné con un requisito formulado en la literatura en torno a las explicaciones reductivas de la conciencia y expuse los argumentos que nos llevan a concluir que la conciencia fenoménica no puede satisfacerlos, fundamentalmente por el tipo de análisis conceptual que no puede hacerse de los estados fenoménicamente conscientes.

Adicionalmente he considerado un par de modelos alternativos de explicación reductiva, según los cuales podrían darse cuenta del caso de la conciencia pues no requieren de análisis

conceptual. Pero he expuesto también una crítica acerca hacia éstos que los señala como no reductivos y con limitaciones respecto a la transparencia epistémica que pueden ofrecer. Sin embargo, aunque la respuesta a estos modelos fuese concluyente, esto no garantiza que sean las únicas alternativas.

Considero que si bien los argumentos dados en esta tesis no son concluyentes (como es de esperarse en filosofía), muestran que hay razones de peso para inclinar la balanza hacia ciertas conclusiones en el debate. Además considero que se han expuesto de forma adecuada tanto la dialéctica como las nociones involucradas en este debate, lo que era también uno de los objetivos generales. Así, además de esta breve recapitulación de lo realizado, en el resto de esta sección de conclusiones expondré algunas posibles vías para continuar con el debate. No profundizaré en ellas pues mi objetivo en esta sección de conclusiones es simplemente mostrar que sé que el debate está lejos de concluirse y plantear algunas líneas de investigación futuras.

Además de cuestionar cada una de las premisas, se pueden atacar algunos de los supuestos. Por ejemplo, si el fisicalismo resulta ser falso y la conciencia algo ontológicamente independiente, explicarla en términos físicos es un camino incorrecto. El problema real sería el de explicar la relación entre dos tipos de entidades ontológicamente distintas.

Otras vías que vale la pena señalar se relacionan con el modelo ND. Si bien he expuesto que la tesis de la implicación *a priori*, sobre la que se discute normalmente en filosofía de la mente, es compatible con el modelo ND de explicación científica, no se ha dicho que esta no sea compatible con otros modelos, aunque eso es algo que debe mostrarse pues tampoco es evidente. De esta forma, existe la posibilidad de que en otros modelos explicativos podamos lograr mayor avance aun aceptando la conclusión de que los hechos fenoménicos no sean implicados *a priori* por P, pues otros modelos podrían no comprometerse con la tesis de la implicación *a priori*.

En relación con lo anterior, cabe mencionar también que el modelo ND se ha propuesto como un modelo explicativo para ciencias más avanzadas –por ejemplo la física- en las que hay mayor número de leyes establecidas. Explicar en este modelo, como se ha señalado, depende de las leyes, hipótesis o teorías ya instituidas, de allí que se llame “nomológico”. En cambio las ciencias cognitivas son de relativo reciente nacimiento y aun no se dispone de muchas leyes, lo que podría ser una limitante para el uso de este modelo. Además, aun aceptando este modelo, se podrían replantear los objetivos y las aspiraciones acerca de qué exactamente queremos explicar respecto a la conciencia, pues algunas de estas cosas podrían explicarse adecuadamente ya sea en los términos expuestos (Carruthers (2005)) u en otros.

Finalmente, este modelo, como ha sido señalado también, ha estado expuesto a muchas críticas. Algunos, por ejemplo Horst (2007), han declarado que este modelo sufre de un (casi) total descrédito en la filosofía de la ciencia desde mediados del siglo XX, por lo cual debatir con base en este modelo, como Horst reconoce que hace la filosofía de la mente contemporánea, resultaría obsoleto. Si bien considero que la conclusión de Horst es algo exagerada, pues el modelo ND podría tener algunas ventajas y no parece estar totalmente abandonado, valdría la pena explorar otros modelos de explicación. (De hecho, valdría la pena explorar otros modelos de explicación reductiva aun cuando el presentado fuese correcto).

Por otro lado, actualmente hay diversos argumentos deflacionistas (a parte de los formulados por Block y Stalnaker) que señalan que la tesis de la implicación *a priori* no es el caso incluso para hechos macroscópicos en los que la conciencia no tiene nada que ver. Esto, argumentan, a partir de que el análisis conceptual requerido no es tal como sus proponentes suponen, sino que es de un tipo demasiado exigente para ser satisfecho. Si esto es el caso, argumentar a partir de la tesis de la implicación *a priori* podría no tener conclusiones (epistémicas u ontológicas) relevantes.

Argumentos de este tipo pueden encontrarse en Díaz-León (2010), Elpidorou (2013) entre otros, quienes señalan que para los propósitos de Chalmers y Jackson no es sólo necesario que un sujeto sea capaz de asignar el referente de un término a partir de una descripción, sino que debe poder hacerlo cuando esta descripción está formulada en términos enteramente microfísicos. No es claro que esto sea parte de las condiciones de entender y poseer un concepto cualquiera. Esta es a mí parecer una de las vías más interesantes y prometedoras del debate, cuyas consecuencias no son sólo a nivel de las explicaciones, sino metafísicas.

Finalmente, dado que esta brecha explicativa es causada en gran medida por el tipo de conceptos que usamos para referir a los estados fenoménicamente conscientes, es posible que mayor reflexión en torno a la naturaleza y semántica de los conceptos fenoménicos pueda arrojar luz a cómo damos cuenta de la brecha explicativa y quizás a acortarla (sin ser esto incompatible con algunas estrategias ya mencionadas). Sin embargo, todas estas vías son complejas y requieren de un tratamiento detallado que exceden los límites de esta tesis.

Referencias:

- Balog, K. (2009). Phenomenal concepts. En Ansgar Beckermann, Brian P. McLaughlin y Sven Walter (Eds.), *Oxford Handbook of Philosophy of Mind*. (pp. 292-312). New York. Oxford University Press.
- Block, N. (1990). Inverted earth. *Philosophical Perspectives*. 4, (n/a). 53-79.
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and brain sciences*. 18. 227-287
- Block, N. y Stalnaker, R. (1999). Conceptual Analysis, Dualism, and the explanatory gap. *Philosophical review*. 108. (1). 1-46
- Carruthers, P. (2000) *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge. Cambridge University Press.
- Carruthers, P. (2005). Reductive explanation and the “explanatory gap”. *Canadian Journal of Philosophy*. 34. (2). 153-174
- Chalmers, D. (1996). *The conscious mind. In Search of a Fundamental Theory*. New York. Oxford University Press.
- Chalmers, D. (2002). On sense and intentions. *Philosophical Perspectives*, 16, Language and Mind. 135-182

- Chalmers, D. (2002a). Does Conceivability entail possibility? En Tamar S. Gendler y John Hawthorne (Eds.). *Conceivability and Possibility*. (pp. 145-200). New York. Oxford University Press.
- Chalmers, D. (2003). Consciousness and its place in nature. En Stephen P. Stich y Ted A. Warfield (Eds.) *Blackwell guide to the philosophy of mind*. (pp. 102-142). Blackwell.
- Chalmers, D. (2004). The representational character of experience. En Brian Leiter (Ed.), *The Future for Philosophy*. (pp. 153-181) New York. Oxford University Press.
- Chalmers, D. (2009). The Two-Dimensional Argument against Materialism . En Ansgar Beckermann, Brian P. McLaughlin y Sven Walter (Eds.), *Oxford Handbook of Philosophy of Mind*. (pp. 313-33). New York. Oxford University Press.
- Chalmers, D. (2010). Appendix: Two-dimensional semantics. En *The Character of Consciousness*. (pp. 541-568). Oxford University Press.
- Chalmers, D. (2013). Panpsychism and Panprotopsyhism En The Amherst Lecture in Philosophy 8 (1–35). Recuperado de <<http://www.amherstlecture.org/chalmers2013/>>
- Chalmers, D. y Jackson, F. (2001). Conceptual Analysis and reductive explanation. *The Philosophical Review*, 110, (3). 315-360
- Churchland, P. (1985). Reduction, Qualia and the Direct Introspection of Brain States. *Journal of Philosophy*. 82. 8-28
- Diaz-León, E. (2011), Reductive explanation, concepts, and a priori entailment. *Philosophical Studies*. 155 (1). 96-116.
- Elpidorou, A. (2013), Blocking the a priori passage. *Acta Analytica*. 29. (3). 285-307
- Fodor, J. (1974). Special sciences (or: the disunity of sciences as a working hypothesis). *Synthese*, 28 (2). 97-115

- Frege, G. (1892 (2016)). "Sobre sentido y referencia" En Frege, G. *Escritos sobre lógica, semántica y filosofía de las matemáticas*. (pp. 249-277). [Comp. Margarita Valdés]. Universidad Nacional Autónoma de México- Instituto de Investigaciones Filosóficas. México.
- Hempel, C.G. (1965). Aspects of scientific explanation. *Journal of Symbolic Logic*. 37. (4). 747-749
- Hempel, C. G. y Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of science*. 15. (2). 135-175
- Hill, C. (1997). Imaginability, Conceivability, possibility, and the mind-body problem. *Philosophical studies*, 87 (1): 61-85
- Horst, S. (2007). The Demise of Reductionism in Philosophy of Science En *Beyond reduction: philosophy of mind and postreductionist philosophy of science*. (pp. 47-64). Oxford University Press.
- Huemer, W. (2018) "Franz Brentano", *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta (Ed.), URL = <<https://plato.stanford.edu/entries/brentano/>>.
- Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*. 32 (April). 127-136
- Jackson, F. (1986). What Mary didn't know. *Journal of Philosophy*. 83 (May). 291-295
- Jackson, F. (1993). Armchair metaphysics. En Michael, M. y O'leary-Hawthorne. *Philosophy in mind, the place of philosophy in the study of mind*. *Philosophical studies*. (60). 23-43
- Kim, J. (1993). Concepts of Supervenience. En *Supervenience and mind*. (pp. 53-78) Cambridge University Press.
- Kim, J. (1998). *Mind in a physical World: an essay on the mind-body problem and mental causation*. Cambridge, Massachusetts. MIT press.
- Kim, J. (2005). *Physicalism or something near enough*. Princeton monographs in philosophy. Princeton university press.

- Kripke, S. (1980 (2017)). *El nombrar y la necesidad*. [Trad. Margarita Valdés]. Universidad Nacional Autónoma De México- Instituto de investigaciones filosóficas.
- Levine, J. (1983). Materialism and qualia: The Explanatory gap. *Pacific philosophical quarterly*. 64. 356-361
- Lewis, D. (1966). An argument for the identity theory. *Journal of Philosophy*. 63. 17–25
- Lewis, D. (1990) *What Experience Teaches*. En William G. Lycan (Ed.), *Mind and Cognition*. (pp. 29-57) Blackwell.
- Loar, B. (1990). Phenomenal States. *Philosophical Perspectives*. 4. 81-108
- Loar, B. (1997). Phenomenal States (second version). En Ned Block, Owen Flanagan y Guven Güzeldere (Eds.), *The Nature of consciousness. Philosophical debates*. (pp.). MIT Press (1997)
- Lycan, W. (1996) *Consciousness and Experience*. Cambridge, Massachusetts. MIT Press.
- McGinn, C. (1991). Can we solve the mind-body problem?. *Mind* .98. (July). 349-66
- Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientific Explanation*.
- Nagel, T. (1974). What Is It Like to Be a Bat?. *The Philosophical Review*, 83 (4), pp. 435-450
- Putnam, H. (1973). The nature of mental states. En W. H. Capitan & D. D. Merrill, (Eds). *Art, Mind, and Religion* (pp. 37-48). Pittsburgh University Press.
- Putnam, H. (1975). The meaning of meaning. *Minnesota Studies in the Philosophy of Science* 7. 131-193
- Robinson, H. (2017). Dualism. *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (Ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/dualism/>
- Rosenthal, D. (1986). Two concepts of consciousness. *Philosophical Studies* 49. (May): 329-59

- Salmon, W. (1989). The First Decade (1948-57) Peace in the Valley (but Some Trouble in the Foothills). En *Four decades of scientific explanation*. University of Pittsburg Press.
- Stoljar, D. (2001) Two Conceptions of the Physical. *Philosophical and Phenomenological Research*. 62. (2). 253-281
- Stoljar, D. (2005). Physicalism and phenomenal concepts. *Mind and Language* 20. (2). 296-302.
- Strevens, M. (2006). "Scientific explanation". En Donald M. Borchert (Ed). *Macmillan Encyclopedia of Philosophy*. Macmillan Reference.
- Tye, M. (1995). A representational theory of pains and their phenomenal character. *Philosophical Perspectives*. 9. 223-39.
- Tye, M. (2009). Representationalist Theories of Consciousness. En Ansgar Beckermann, Brian P. McLaughlin y Sven Walter (Eds.), *Oxford Handbook of Philosophy of Mind*. (pp. 253-267). New York. Oxford University Press