

TESIS DOCTORAL

2015

The logo of the Universidad Nacional de Educación a Distancia (UNED), consisting of the letters 'UNED' in white on a dark green square background.

---

**LIBRE ALBEDRÍO Y RESPONSABILIDAD  
MORAL: UNA APROXIMACIÓN DESDE LA  
NEUROCIENCIA**

José Manuel Muñoz Ortega

Licenciado en Biología

---

Departamento de Lógica,  
Historia y Filosofía de la Ciencia

Facultad de Filosofía

Director:  
Dr. D. Carlos J. Moya Espí

Tutor:  
Dr. D. David Teira Serrano

UNED (Universidad Nacional de Educación a Distancia)

Departamento de Lógica, Historia y Filosofía de la Ciencia

Facultad de Filosofía

Libre albedrío y responsabilidad moral:  
una aproximación desde la neurociencia

Tesis doctoral

Autor: José Manuel Muñoz Ortega (Licenciado en Biología)

Director: Dr. D. Carlos J. Moya Espí (Universidad de Valencia)

Tutor: Dr. D. David Teira Serrano (UNED)

Programa de doctorado: Lógica, Historia y Filosofía de la Ciencia



## Agradecimientos

Son varias las personas a las que quiero expresar mi breve pero sincero agradecimiento por su ayuda durante la elaboración de esta tesis doctoral.

Con Carlos Moya estaré siempre en deuda por haber aceptado dirigir mi tesis doctoral. Desde el principio puso todo su conocimiento y su experiencia a mi disposición, revisando el manuscrito y efectuando valiosísimas observaciones que a buen seguro han servido para mejorar ostensiblemente el trabajo. También me brindó palabras de ánimo a pesar de los altibajos que se produjeron durante el estudio y la escritura. Por si fuera poco, todo ello fue mostrando siempre una humildad que, creo, es doblemente virtuosa teniendo en cuenta que se trata de un especialista de reconocido prestigio internacional.

A David Teira le debo un interminable “gracias” por el igualmente interminable número de correos electrónicos a los que me ha contestado a lo largo de estos años, siempre con amabilidad y paciencia. También le agradezco sinceramente tanto sus ánimos como su disposición a hacer siempre de los trámites académicos una carga lo más ligera posible. Y, por supuesto, no olvido su generosa ayuda a la hora de buscar dirección para mi tesis.

La investigación predoctoral que ha resultado en esta tesis ha conducido a tres publicaciones académicas de mi autoría, las cuales, con adaptaciones muy menores, han sido incluidas aquí con la amable autorización de las revistas correspondientes. Dichas publicaciones, junto con las partes de la tesis donde principalmente se incluyen, se listan a continuación:

Muñoz, J.M.. 2012. “Hacia una sistematización de la relación entre determinismo y libertad”. *Daimon: Revista Internacional de Filosofía* 56: 5-19. → Capítulo 1.

Muñoz, J.M. 2013. “Neurofilosofía y libre albedrío”. *Daimon: Revista Internacional de Filosofía* 59: 57-70. → Capítulo 3.

Muñoz, J.M. 2015. “Mecánica cuántica y libre albedrío: cinco cuestiones fundamentales”. *Principia: Revista Internacional de Epistemología* 19(1): 65-92. → Capítulo 2.

Pero es a mi familia a quien corresponde mi mayor agradecimiento. Son muchos los momentos que he dejado de disfrutar y compartir con ellos por dedicarme a la tarea que esta tesis suponía. Gracias a mi padres, M<sup>a</sup> Salud y José Carlos, que siempre me han dado su amoroso apoyo en todo cuanto tuviera que ver con mi formación. Gracias a mi hermano, Carlos, porque con su particular sentido del humor ha contribuido a aliviar los momentos de desánimo que inevitablemente han surgido. Y gracias a Daniela, Miguel Ángel y Luna; sin su cariño y su paciencia no habría podido hallar la paz de espíritu que me ha acompañado durante mi trabajo bajo el techo del hogar que muy felizmente compartimos.

En Mahón, noviembre de 2015.

# Índice

|  |    |
|--|----|
| <i>Agradecimientos</i> .....   | 3  |
| <i>Lista de figuras</i> .....  | 9  |
| INTRODUCCIÓN .....   | 11 |
| CAPÍTULO 1. LA ESTRUCTURA DE LOS PROBLEMAS SOBRE EL<br>LIBRE ALBEDRÍO Y LA RESPONSABILIDAD MORAL ..... | 19 |
| 1.1. Premisas para la sistematización .....  | 20 |
| 1.2. Compatibilismo .....  | 22 |
| 1.3. Incompatibilismo .....  | 31 |
| 1.4. Posturas rupturistas .....  | 35 |
| CAPÍTULO 2. EXPLORANDO EL NIVEL “MICRO”: MECÁNICA<br>CUÁNTICA Y LIBRE ALBEDRÍO .....                   | 39 |
| 2.1. Aspectos fundamentales de la mecánica cuántica. Interpretación<br>de Copenhague .....             | 40 |
| 2.1.1. Descubrimiento de los cuantos .....   | 40 |
| 2.1.2. Dualidad onda-partícula .....   | 42 |
| 2.1.3. No-localidad .....  | 44 |
| 2.1.4. Principio de incertidumbre .....  | 44 |
| 2.1.5. Ondas y probabilidad .....  | 46 |
| 2.1.6. Problema de la medida .....   | 48 |
| 2.1.7. Desafío al determinismo .....   | 50 |

|   |    |
|---|----|
| 2.2. Interpretaciones alternativas y determinismo .....                                     | 52 |
| 2.2.1. La suma de historias posibles y los mundos múltiples .....                           | 53 |
| 2.2.2. La interpretación de Bohm .....  | 54 |
| 2.3. ¿Existe una correlación entre el indeterminismo cuántico y el<br>libre albedrío? ..... | 55 |
| 2.3.1. Posturas a favor .....   | 55 |
| 2.3.2. Posturas en contra .....   | 58 |
| 2.4. ¿Qué dice la neurobiología? .....  | 60 |
| 2.5. El caos como amplificación .....   | 65 |
| 2.6. Mecánica cuántica y libre albedrío: análisis crítico .....                             | 66 |
| 2.6.1. Irrelevancia para la compatibilidad y la significatividad .....                      | 67 |
| 2.6.2. El control último como reto para la inteligibilidad .....                            | 68 |
| 2.6.3. Indicios contra la existencia .....  | 72 |
| 2.6.4. Dudas sobre el determinismo .....  | 75 |

### CAPÍTULO 3. EXPLORANDO EL NIVEL “MACRO”:

|  |     |
|--|-----|
| NEUROFILOSOFÍA Y LIBRE ALBEDRÍO .....  | 81  |
| 3.1. La neurofilosofía y la búsqueda de un nexo entre el cerebro<br>y la mente ..... | 82  |
| 3.2. El enfoque de la neurofilosofía mínima .....                                    | 85  |
| 3.3. Bases biológicas de la volición .....   | 89  |
| 3.3.1. Un poco de neuroanatomía .....  | 89  |
| 3.3.2. Estudios de lesiones .....  | 93  |
| 3.3.3. Estudios de neuroimagen funcional .....                                       | 96  |
| 3.3.4. Estudios de neuroelectrofisiología .....                                      | 98  |
| 3.3.5. Algunos comentarios adicionales .....   | 100 |
| 3.4. Benjamin Libet y la génesis temporal de actos voluntarios .....                 | 103 |
| 3.5. Daniel Wegner y la ilusión de la voluntad consciente .....                      | 112 |
| 3.6. Neurofilosofía y causalidad descendente .....                                   | 114 |

|  |     |
|--|-----|
| 3.7. Neurofilosofía y libre albedrío: análisis crítico ..... | 118 |
| 3.7.1. Neutralidad sobre la compatibilidad .....             | 119 |
| 3.7.2. Relevancia parcial sobre la significatividad .....    | 121 |
| 3.7.3. Dudas sobre la inteligibilidad .....                  | 123 |
| 3.7.4. Dificultades para aceptar la existencia .....         | 126 |
| 3.7.5. Inconclusión sobre el determinismo .....              | 128 |

CAPÍTULO 4. CAUSALIDAD MENTAL, INDETERMINISMO Y CONTROL ÚLTIMO: ¿ES POSIBLE UN LIBERTARISMO EN ARMONÍA CON LA CIENCIA? .....

|   |     |
|---|-----|
| 4.1. Primera condición: causalidad mental .....   | 132 |
| 4.1.1. Argumentos en contra de la causalidad mental .....   | 134 |
| 4.1.2. Causalidad intencional: ¿existe un vínculo causal entre razones y acción? .....                  | 142 |
| 4.1.3. Papel de los estados fenomenológicos en la causalidad mental .....                               | 146 |
| 4.2. Segunda condición: correlación inteligible entre la causalidad mental y un sustrato material ..... | 152 |
| 4.2.1. Causalidad mental y materialismo: en busca de una integración .....                              | 153 |
| 4.2.2. Reduccionismo, superveniencia y emergencia .....   | 158 |
| 4.2.3. Sistemas complejos y causalidad descendente .....  | 163 |
| 4.3. Tercera condición: indeterminación ontológica que permita el control último .....                  | 171 |
| 4.3.1. Indeterminismo endógeno: la propuesta de Kane .....  | 172 |
| 4.3.2. Indeterminismo exógeno: la propuesta de Moya .....   | 176 |

CAPÍTULO 5. LIBRE ALBEDRÍO Y NEUROCIENCIA: UNA PROPUESTA LIBERTARISTA .....

|                              |     |
|------------------------------|-----|
| 5.1. La poda sináptica ..... | 182 |
|------------------------------|-----|

|  |         |
|--|---------|
| 5.2. La “poda semántica” como modalidad de causalidad mental .....   | 187     |
| 5.3. Indeterminismo y causalidad mental: las vías exógena y<br>endógena .....  | 193     |
| 5.4. El Grupo Bruselas-Austin y los sistemas alejados del equilibrio<br>termodinámico .....  | 195     |
| 5.5. Hofstadter y los bucles extraños .....  | 198     |
| 5.6. La “poda deliberativa” como modalidad de causalidad mental ....   | 210     |
| 5.7. Observaciones sobre la poda deliberativa (I): causalidad<br>mental .....  | 226     |
| 5.8. Observaciones sobre la poda deliberativa (II): correlación<br>inteligible entre la causalidad mental y un sustrato material ..... | 234     |
| 5.9. Observaciones sobre la poda deliberativa (III): indeterminación<br>ontológica que permita el control último .....                 | 239     |
| <br>CONCLUSIONES .....   | <br>255 |
| <br><i>Bibliografía</i> .....  | <br>269 |

## Lista de figuras

|   |     |
|---|-----|
| 2.1. Experimento de la doble rendija .....  | 43  |
| 2.2. Estructura básica de una neurona .....   | 61  |
| 2.3. Membrana celular con canal iónico de sodio .....   | 63  |
|   |     |
| 3.1. Encéfalo humano con cerebelo y lóbulos del córtex cerebral .....   | 90  |
| 3.2. Córtex prefrontal en el ser humano y en el macaco .....  | 91  |
| 3.3. Ubicación del cerebelo y de algunas estructuras subcorticales .....  | 92  |
|   |     |
| 5.1. Proceso de poda sináptica .....  | 183 |
| 5.2. Relación entre edad y densidad de espinas dendríticas en el<br>córtex prefrontal dorsolateral .....  | 185 |
| 5.3. Patrones de realimentación de vídeo obtenidos por Hofstadter .....   | 199 |
| 5.4. Patrones dinámicos con forma de espiral obtenidos por Hofstadter .....   | 200 |
| 5.5. <i>Waterfall</i> , de M. C. Escher .....   | 201 |
| 5.6. <i>Drawing hands</i> , de M. C. Escher .....   | 204 |
| 5.7. Ejemplo de cambios temporales en la probabilidad que distintas<br>posibilidades alternativas tienen de convertirse en decisión como<br>resultado de una deliberación iterativa ..... | 217 |
| 5.8. Diferentes combinaciones de tiempo de deliberación y contenido de<br>la decisión de Alice (personaje ideado por van Inwagen) .....   | 253 |



# Introducción

«El que haya impedido que pasara no cambia el hecho de que fuera a pasar.»

*Minority Report* (Steven Spielberg, 2002).<sup>1</sup>

La libertad es algo que nos incumbe a todos. Desde que el hombre es hombre, probablemente la pregunta más importante que se ha hecho a sí mismo es la de si es un ser libre o, por el contrario, un barco a la deriva en el océano del universo, un capricho más de la naturaleza sujeto al inevitable curso del destino. Hay muchos tipos de libertad: para hablar, para creer, para actuar... para comprar ese disco de mi grupo favorito, para viajar a Venecia en mi luna de miel. Pero quizá todas esas libertades, algunas más superficiales que otras, sean reducibles a un solo tipo de libertad, enraizado en un nivel más profundo. Porque, ¿qué valor tendría mi libertad para hablar, concebida como derecho a decir unas palabras en un momento y un lugar determinados, si no fuera *mi voluntad* decir esas palabras en ese momento y en ese lugar? Una libertad para hablar sin el soporte de mi voluntad sería sólo *libertad* desde un punto de vista superficial.<sup>2</sup> Por ello, quizá no tenga sentido analizar los orígenes de los diferentes tipos de libertad, desde un punto de vista metafísico, si no es

---

<sup>1</sup> El guión de esta película es obra de Scott Frank y Jon Cohen, y está basado en una historia escrita por Philip K. Dick. El fragmento transcrito corresponde a la versión doblada al castellano.

<sup>2</sup> Kane (2005, pp. 2-4) ofrece una completa explicación de las diferencias entre lo que él llama “libertades superficiales” y el libre albedrío entendido como una capacidad más profunda que estas.

basándonos en una voluntad última que los sustente: en mí, en su, en nuestro *libre albedrío*. Nadie sabe con certeza si esta capacidad existe o no realmente. El presente trabajo pretende abordar el amplio debate alrededor de este problema fundamental desde una perspectiva basada en la ciencia contemporánea.

La relación entre el libre albedrío y el determinismo es, no cabe duda, uno de los problemas más ampliamente debatidos en la historia de la filosofía. Y es lógico que así sea si tenemos en cuenta sus profundas implicaciones en diversos ámbitos, no solo filosóficos, sino también científicos y religiosos. ¿Estamos inevitablemente determinados por las condiciones de nuestro nacimiento, nuestra educación y nuestro entorno social? ¿Serían nuestras acciones predecibles por parte de alguien que conociera por completo nuestras características psicológicas? ¿Somos máquinas biológicas completamente programadas por nuestros genes para pensar y actuar tal y como a estas diminutas entidades biológicas les conviene que hagamos para su propagación? Incluso, ¿estamos predestinados por la voluntad de un ente superior y todopoderoso que dirige el transcurrir de los acontecimientos? Como vemos, las implicaciones de la relación entre libertad y necesidad, entre libre albedrío y determinismo, son de naturaleza ética, jurídica, social, política, psicológica, biológica, metafísica, religiosa...

Todas las preguntas recién expuestas dependen en gran medida de qué tipo de determinismo sea el que consideremos. Así, según el contexto en que nos situemos, podemos hablar de determinismo físico, biológico, psicológico, social, teológico... No obstante, y a efectos prácticos, es posible reducir todas estas variantes, a excepción del determinismo teológico, a una sola. He aquí una breve pero útil definición de determinismo ofrecida por Fischer *et al.* (2007, p. 2): “[P]odemos tratar el determinismo como la tesis según la cual en cualquier instante [...] el universo posee exactamente un solo futuro posible físicamente”. Peter van Inwagen (1983, p. 65) ofrece una conocida descripción, más completa que la anterior, según la cual el determinismo es la conjunción de dos tesis: (1) “Para cada instante de tiempo, existe una proposición que expresa

el estado del mundo en ese instante”, y (2) “[s]i  $p$  y  $q$  son proposiciones cualesquiera que expresan el estado del mundo en algunos instantes, entonces la conjunción de  $p$  junto con las leyes de la naturaleza implica  $q$ ”. En la misma obra y página, el autor insiste en que “[e]l determinismo es, intuitivamente, la tesis según la cual, dados el pasado y las leyes de la naturaleza, existe sólo un futuro posible”.

Como ya avanzábamos, estas definiciones no son aplicables al caso del determinismo teológico, el cual no se basa en leyes naturales conocidas por los seres humanos y, por lo tanto, puede ser entendido como la voluntad de un ser superior cuyas propias leyes y/o designios escapan a nuestro entendimiento. Es por ello que cualquier estudio del libre albedrío puede (y debería) ser abordado desde la perspectiva de dos relaciones muy diferentes: libre albedrío y determinismo “natural” (al que bautizo de esta forma porque las leyes naturales son elemento nuclear del mismo), o libre albedrío y determinismo teológico. Mientras que la segunda merece un estudio pormenorizado y especializado que no abordaremos aquí, la primera de estas relaciones constituye el marco del presente trabajo.

En esta tesis doctoral estudiamos el problema de la relación entre el libre albedrío y el determinismo desde una perspectiva basada en los conocimientos aportados por la ciencia contemporánea, con un especial hincapié en la neurociencia, y elaboramos al respecto una propuesta propia de índole libertarista. El trabajo está estructurado en cinco capítulos:

1. La estructura de los problemas sobre el libre albedrío y la responsabilidad moral.
2. Explorando el nivel “micro”: mecánica cuántica y libre albedrío.
3. Explorando el nivel “macro”: neurofilosofía y libre albedrío.
4. Causalidad mental, indeterminismo y control último: ¿es posible un libertarismo en armonía con la ciencia?
5. Libre albedrío y neurociencia: una propuesta libertarista.

En el capítulo 1 situamos el marco general del debate sobre la relación entre libre albedrío y determinismo, y proponemos una organización sistemática de las principales posiciones acerca de dicha relación. La sistematización se basa en cuatro premisas: (a) el determinismo impide la existencia de posibilidades alternativas de decisión y/o acción, (b) las posibilidades alternativas son necesarias para decidir y/o actuar libremente, (c) el determinismo impide el control último sobre nuestras decisiones y/o acciones, y (d) el control último es necesario para decidir y/o actuar libremente. Diversas combinaciones de la aceptación/rechazo de estas premisas conducen a diferentes teorías: compatibilismo, libertarismo, determinismo duro, incompatibilismo duro y algunas propuestas relacionadas con este último. No obstante, también presentamos ciertas posiciones para las cuales no son de aplicación las premisas mencionadas.

En los capítulos 2 y 3 llevamos a cabo un estudio crítico sobre la relación existente entre los conocimientos científicos contemporáneos y el problema del libre albedrío. Dicho estudio se efectúa a dos escalas: la mecánico-cuántica o microscópica (capítulo 2) y la neural o macroscópica (capítulo 3). La estructura interna de ambos capítulos es similar: en primer lugar, se exponen los conocimientos científicos que resultan fundamentales para el estudio, y, en segundo lugar, se analiza su relevancia para la relación entre determinismo y libre albedrío. El análisis se efectúa desde ángulos diversos, representados por cinco cuestiones fundamentales propuestas por Robert Kane, de las cuales las cuatro primeras surgen en su obra *The Significance of Free Will* (Kane 1996, p. 13), mientras que la quinta surge más tarde (Kane 2002a, p. 6):

[A] La Cuestión de Compatibilidad: ¿Es el libre albedrío compatible con el determinismo?

[B] La Cuestión de Significatividad: ¿Por qué queremos, o deberíamos querer, poseer un libre albedrío que es incompatible con el determinismo? ¿Es este un tipo de libertad “que valga la pena querer” (por utilizar la útil frase de Dennett)? Y si es así, ¿por qué?

[C] La Cuestión de Inteligibilidad: ¿Podemos darle sentido a una libertad o libre albedrío que es incompatible con el determinismo? ¿Es dicha libertad coherente o inteligible? ¿O es, como reivindican muchos críticos, esencialmente misteriosa y terminantemente oscura?

[D] La Cuestión de Existencia: ¿Dicha libertad existe en realidad en el orden natural? Y si es así, ¿dónde?

[E] La Cuestión Determinista: ¿Es el determinismo cierto?<sup>3</sup>

En el capítulo 4 estudiamos las posibilidades del libertarismo de dar con una teoría del libre albedrío que esté en armonía con los conocimientos aportados por la ciencia contemporánea. A nuestro juicio, dicha teoría debería pasar por el cumplimiento de, al menos, tres condiciones fundamentales:

- 1) *Causalidad mental*. Para ejercer nuestra libertad resulta especialmente importante la causalidad intencional: la causalidad mental de los estados intencionales.
- 2) *Correlación inteligible entre la causalidad mental y un sustrato material*. Dicha correlación no puede conducirnos al epifenomenismo, sino que debe traer consigo un sentido descendente de la causalidad.
- 3) *Indeterminación ontológica que permita el control último*. Dicha indeterminación, desde luego, debe resistir al argumento de *Mind*, según el cual el indeterminismo excluye el control del agente sobre sus decisiones y/o acciones porque conlleva el azar y no existe control sobre aquello que es azaroso.

A lo largo del capítulo analizamos en profundidad cada una de estas condiciones, lo que nos proporciona la base para presentar una propuesta propia en el capítulo 5. Con esta, que podemos enmarcar en la corriente del materialismo no reductivo, tratamos de aliviar la dificultad que supone para el

---

<sup>3</sup> Ofrecer una respuesta negativa a dicha cuestión sería equivalente a afirmar que *no* es cierto que el pasado y las leyes de la naturaleza nos conduzcan *siempre* hacia un único futuro posible. Con este fin, sería necesario describir, al menos, *un suceso natural* que no funcionase de modo determinista, de forma que el determinismo no sería cierto como ley universal (si bien podría ser cierto como ley aplicable a casos particulares, pues seguiría habiendo sectores de la realidad funcionando de modo determinista).

libertarismo la Cuestión de Inteligibilidad (mírese más arriba).<sup>4</sup> La propuesta está influenciada por los intentos de naturalizar lo mental desarrollados por Dretske (1988, 1993, 2009), Van Gulick (1993) y Moya (2011, pp. 197-200), y busca compatibilizar estos con las siguientes herramientas conceptuales:

- a. *Emergencia contextual*, en la que “[l]a descripción de propiedades en un nivel de descripción particular ofrece condiciones *necesarias pero no suficientes* para obtener la descripción de propiedades en un nivel superior” (Bishop y Atmanspacher 2006, p. 1757).
- b. *Sistemas no integrables*, cuya descripción se realiza a través de distribuciones, y no por medio de puntos u objetos particulares. Dichas distribuciones son irreducibles a las trayectorias individuales, y su estudio es de carácter estadístico.
- c. *Bucles extraños*, en los que un ascenso o un descenso por una sucesión de niveles jerárquicos acaba conduciendo al nivel desde el cual se comenzó. En otras palabras, “un bucle extraño es un bucle de realimentación paradójico con pasos de nivel” (Hofstadter 2007, pp. 101-2).
- d. *Poda sináptica*, mecanismo neural en virtud del cual: 1) disminuye el número de axones que van a parar a una misma célula postsináptica y también el número de células de este tipo a las que va a parar un mismo axón, y 2) se refuerzan los vínculos intercelulares supervivientes mediante la formación de sinapsis adicionales.

El resultado final de nuestra propuesta queda plasmado en una tesis que denominamos *poda deliberativa*, y según la cual la imposición de constricciones semánticas con estructura de bucle extraño por parte del contexto normativo interno en la mente del agente volitivo desencadena un proceso de poda sináptica en el que se seleccionan y refuerzan rutas neurales conducentes a la

---

<sup>4</sup> Nos centraremos únicamente en la mencionada Cuestión de Inteligibilidad, y no en las otras cuatro cuestiones de Kane: de Compatibilidad, de Significatividad, de Existencia y Determinista.

ejecución de ciertas acciones. Y ello en el seno de un sistema no integrable (el cerebro) que está englobado en un sistema complejo (el ser humano), y en virtud de una relación de emergencia contextual entre las propiedades mentales y las que son propias de niveles jerárquicos inferiores al mental. Con nuestra propuesta, que consideramos novedosa en tanto en cuanto integra un mecanismo neurobiológico muy específico y no empleado anteriormente para defender el libre albedrío, pretendemos poner nuestro pequeño grano de arena a la hora de intentar explicar de qué modo podrían nuestros estados mentales tener una influencia decisiva, indeterminada e inteligible sobre el mundo físico.

Antes de entrar en materia, nos gustaría añadir dos breves aclaraciones metodológicas. En primer lugar, dada la naturaleza de la investigación que ha resultado en esta tesis, a la imprescindible bibliografía filosófica utilizada se ha sumado un importante número de obras de índole científica en general, y neurocientífica en particular. En segundo lugar, para formatear la bibliografía hemos tomado como modelo el estilo autor-fecha de Kate L. Turabian, que es el empleado por *Theoria*, la principal revista editada actualmente en nuestro país en la especialidad de filosofía de la ciencia.



# Capítulo 1. La estructura de los problemas sobre el libre albedrío y la responsabilidad moral

«¡No, no, muchacho, deja eso en nuestras manos!  
¡Puedes estar contento!  
¡En un par de semanas serás un hombre libre!»

*La naranja mecánica* (Stanley Kubrick, 1971).<sup>1</sup>

En el presente capítulo abordamos una sistematización de las discusiones sobre la relación entre el libre albedrío y el determinismo. Alrededor de un debate tan antiguo, amplio y en ocasiones confuso, han aparecido posiciones extremadamente diversas, razón por la cual se hace necesaria una organización de las principales posturas filosóficas vigentes a día de hoy, encaminada a aportar una perspectiva global. En este sentido son enormemente valiosas aportaciones como la de Moya (2006b), quien ha contribuido a la organización sistemática de los argumentos escépticos sobre la responsabilidad moral, estrechamente relacionados con el problema mencionado, si bien en la sistematización que aquí proponemos las cuestiones acerca de la responsabilidad moral serán tratadas conjuntamente con las cuestiones acerca del libre albedrío. Este enfoque, que es corriente en la literatura sobre el tema, será, dicho sea de paso, el que mantendremos a lo largo de todo el trabajo. No

---

<sup>1</sup> El guión de esta película, también obra de Kubrick, se basó en la novela homónima de Anthony Burgess. El fragmento transcrito corresponde a la versión doblada al castellano.

obstante, debemos aclarar que no está adoptado por unanimidad. Algunos autores se han desmarcado sosteniendo el denominado “argumento directo”, en el cual se debate la compatibilidad/incompatibilidad entre determinismo y responsabilidad moral sin considerar previamente la compatibilidad/incompatibilidad entre determinismo y libertad. Efectuaremos, por ello, la oportuna aclaración cuando dichos autores aparezcan en el texto.

### **1.1. Premisas para la sistematización**

El conflicto entre libertad y determinismo puede parecer obvio a primera vista. Nos vemos a nosotros mismos como agentes libres capaces de modificar nuestro entorno de diversas formas, y pensamos que para ello disponemos de diversas opciones cuyo poder de elección nos corresponde (Kane 2002a, p. 5). Y esta creencia humana (acertada o no) nos da la clave para entender las dos características fundamentales de lo que, desde un punto de vista preliminar, podríamos entender como libre albedrío, a saber, que como ser humano: (1) dispongo de un número, mayor que uno, de posibilidades entre las cuales escoger en mis decisiones y/o acciones, y (2) soy la fuente, el autor, el origen último de estas decisiones y/o acciones. Nuestro concepto preliminar de libre albedrío se basaría, en consecuencia, en dos condiciones fundamentales: (1) posibilidades alternativas de decisión y/o acción, y (2) control último sobre nuestras decisiones y/o acciones.

Según este análisis, es fácil comprender de qué manera podría el determinismo ser una amenaza para la libertad. Si el determinismo es cierto, no tendríamos de posibilidades alternativas sobre las que escoger, pues en cada

instante de tiempo existiría un único futuro posible.<sup>2</sup> Además, no podríamos ser el origen último de nuestras decisiones y/o acciones, sino que esta función recaería en la conjunción de los eventos pasados y las leyes de la naturaleza. Sin embargo, la incompatibilidad aparente entre necesidad y libertad, entre determinismo y libre albedrío, ha sido motivo de una renovada y amplia discusión, a lo largo de las últimas décadas, entre los partidarios y detractores de esta incompatibilidad. La sistematización que aquí proponemos para situar ordenadamente dicha discusión utilizará como punto de partida cuatro premisas (cf. Kane 2002a, pp. 10-11; cf. Moya 2006b, p. 6):

- (a) El determinismo impide la existencia de posibilidades alternativas de decisión y/o acción.
- (b) Las posibilidades alternativas son necesarias para decidir y/o actuar libremente.
- (c) El determinismo impide el control último sobre nuestras decisiones y/o acciones.
- (d) El control último es necesario para decidir y/o actuar libremente.

Debemos destacar que las discusiones que siguen a continuación carecerían de sentido sin la existencia de otra premisa, de carácter preliminar, a la que llamaremos premisa ( $\alpha$ ): el determinismo es cierto o no lo es. Sin esta verdad lógica carecería de fundamento cualquier teoría compatibilista o incompatibilista. No obstante, cabe puntualizar que ( $\alpha$ ) depende de que el determinismo sea una tesis que goce de contenido definitivo y evaluable, como efectivamente es el caso (Moya 2006b, pp. 10-13).

Diferentes combinaciones de las premisas (a), (b), (c) y (d), según su aceptación o negación, dan lugar a diversas posturas acerca de la relación entre determinismo y libertad. Veámoslas a continuación.

---

<sup>2</sup> Este es el denominado “argumento de consecuencia” o “argumento modal” (véase sección 1.2.).

## 1.2. Compatibilismo

Tal y como hemos visto hasta ahora, la primera pregunta que surge cuando pensamos en la relación del libre albedrío con el determinismo es si son compatibles, es decir, si el determinismo no supone una amenaza para nuestra libertad o si, por el contrario, se trata de una constrictión que, de modo definitivo, nos impide ser capaces de tomar decisiones y realizar acciones con libertad, así como ser dignos de alabanza o reproche por nuestras decisiones y acciones morales.

Son muchos los autores que han argumentado a favor del compatibilismo. Se trata de una posición atractiva por cuanto supone una escapatoria accesible al conflicto entre el determinismo y el libre albedrío, consistente precisamente en la negación de la existencia de dicho conflicto. Desde autores del siglo XVII como Thomas Hobbes hasta otros en la actualidad, se han ofrecido multitud de explicaciones compatibilistas.

Hemos visto anteriormente un concepto preliminar de libre albedrío basado en dos condiciones: (1) posibilidades alternativas de decisión y/o acción, y (2) control último sobre nuestras decisiones y/o acciones. La mayoría de las teorías compatibilistas ofrecidas hasta la fecha se han centrado en un análisis crítico de estas condiciones.

La incompatibilidad entre determinismo y libertad podría ser presentada como conclusión de las premisas (a) y (b), centradas en las posibilidades alternativas de decisión y/o acción, y según las cuales: (a) el determinismo impide la existencia de posibilidades alternativas de decisión y/o acción, y (b) las posibilidades alternativas son necesarias para decidir y/o actuar libremente. Si aceptamos (a) y (b) obtendremos como resultado que el determinismo impide las decisiones y/o acciones libres. No obstante, una gran parte de las explicaciones compatibilistas se ha basado en la negación de una de estas dos premisas.

Un conocido argumento a favor de *(a)* es el llamado “argumento de consecuencia” o “argumento modal”, expuesto, entre otros, por autores como van Inwagen (1975, 1983), Lamb (1977), Widerker (1987), Ginet (1990), Warfield (1996), o Finch y Warfield (1998). Una descripción informal ampliamente debatida e influyente de este argumento es la que ofrece Peter van Inwagen (1983, pp. 16 y 56):

Si el determinismo es cierto, entonces nuestros actos son las consecuencias de las leyes de la naturaleza y los eventos en el pasado remoto. Pero no depende de nosotros lo que sucedió antes de que nacióramos, ni depende de nosotros cuáles son las leyes de la naturaleza. Por lo tanto, las consecuencias de dichas cosas (incluyendo nuestros actos presentes) no dependen de nosotros.

Según este argumento, nada podríamos hacer para cambiar el pasado ni las leyes de la naturaleza y, si aceptamos que el determinismo es cierto, nuestros actos actuales serían consecuencia del pasado y las leyes de la naturaleza, de lo cual se deduce, si asumimos una transferencia de la impotencia, que *no podríamos* hacer nada para que nuestros actos presentes fueran diferentes de lo que son.

La interpretación del significado del verbo “poder” ha sido precisamente el punto en el que han basado sus argumentos algunos autores: los partidarios del llamado “compatibilismo clásico”, iniciado por filósofos como Thomas Hobbes, David Hume o John Stuart Mill. Para los compatibilistas clásicos, *ser libre* significa poseer la capacidad para hacer lo que queremos, deseamos o escogemos hacer, y que no haya nada que nos impida que ejerzamos dicha capacidad. Según esta interpretación, *poder* hacer algo significa gozar de la facultad o la habilidad para hacerlo. Si, pongamos por caso, esta noche voy a cenar con mis amigos, es porque me apetece hacerlo y porque ningún ladrón ha entrado en mi casa y me ha maniatado y amordazado, impidiendo que *pueda* acudir a la cita. ¿Qué sucede con las posibilidades alternativas? La respuesta de los compatibilistas clásicos consiste en que, como agentes libres, somos capaces de hacer algo diferente de lo que hacemos si

gozamos del poder o la habilidad para evitar hacer lo que inicialmente teníamos previsto, y si no existe nada que nos impida que lo hagamos. Si, en lugar de acudir a la cena con mis amigos, decido ir al cine con mi hermano para ver el último film de mi director favorito, es porque mi hermano puede y quiere acompañarme, porque efectivamente ese día estrenan la película que quiero ver y porque ninguno de los amigos a los que iba a encontrarme en la cena ha venido a mi casa y me ha llevado a esta de mala gana. Es importante destacar que, para un compatibilista clásico, el hecho de que yo carezca de impedimentos para hacer algo no significa que de hecho lo vaya a hacer. Lo que ocurre es que, en ausencia de constricciones que me lo impidan, haré algo *si* deseo o escojo hacerlo. En definitiva, los compatibilistas clásicos rechazan la incompatibilidad de la libertad con el determinismo mediante la negación de la premisa (a), que afirma que el determinismo impide la existencia de posibilidades alternativas de decisión y/o acción.

Si aceptamos que *ser libre* no es otra cosa que poder hacer lo que queremos, sin que nada ni nadie nos lo impida, resulta evidente que, aunque nuestra voluntad esté determinada, ello no impide la alternativa de haber hecho algo distinto de lo que hemos hecho si así lo hubiéramos querido. Kane (2002a, p. 13) escribe al respecto:

[S]i el poder para hacer otra cosa significa solamente que habrías hecho otra cosa si hubieras querido o deseado, sería consistente con el determinismo. Podría ser cierto que habrías hecho otra cosa si hubieras querido, aunque esté determinado que de hecho no querías otra cosa. Igualmente, si el poder para hacer otra cosa tiene solo tal significado condicional, podría no requerir cambiar el pasado o violar las leyes de la naturaleza. Decir “podrías haber hecho otra cosa” solamente equivaldría a la afirmación contrafactual de que habrías hecho otra cosa [...] si (en contra del hecho) el pasado [...] o las leyes [...] hubieran sido diferentes de alguna manera, por ejemplo, si hubieras querido o deseado o elegido otra cosa.

Contra este argumento, varios autores (entre los que encontramos, incluso, a compatibilistas como Harry Frankfurt) han alertado del hecho de que aceptarlo supondría apostar por un significado superficial de libertad, basado en la

libertad de acción, pero no de la voluntad en la que dicha acción tiene su sustento.

El compatibilismo clásico niega la premisa (a) de la incompatibilidad entre determinismo y libertad. Otras visiones compatibilistas se basan, en cambio, en la negación de (b): rechazan que las posibilidades alternativas sean necesarias para decidir y/o actuar libremente. Tal es el caso de tres tipos de argumentaciones: los ejemplos basados en personajes, los ejemplos de autocaptura y los ejemplos estilo Frankfurt.

Los ejemplos basados en personajes niegan que las posibilidades alternativas sean necesarias para decidir y/o actuar libremente mediante la exposición de casos de personalidades importantes en la historia de la humanidad desde los puntos de vista ético y moral. Dennett (1984, pp. 131-8) plantea el caso de Martín Lutero. Este afirmó en su día que no podía actuar de forma diferente a como lo hacía, refiriéndose a que su conciencia le impedía actuar de un modo distinto a como lo estaba haciendo. Si aceptamos esto, según Dennett, los actos de Lutero habrían estado determinados por su carácter y sus motivos, pero ello no le privaría de su responsabilidad moral. Más bien al contrario: Lutero con su afirmación habría hecho un alegato en favor de su responsabilidad por todos los actos realizados, que habrían nacido de su libertad de conciencia. Por lo tanto, las posibilidades alternativas no son necesarias para que el hombre sea moralmente responsable. Estas afirmaciones de Dennett siguen la línea iniciada por David Hume, quien opinaba que una persona no puede ser responsable de sus acciones si estas no están determinadas en gran medida por su carácter.

Los ejemplos de autocaptura, o “self-trapping cases” en inglés (Moya 2006b, p. 73), son situaciones en las que un agente se sitúa en una posición en la cual carece de posibilidades alternativas con respecto a una decisión o acción y, sin embargo, se le puede considerar como responsable de dicha decisión o acción. Los ejemplos más conocidos de este tipo han sido ofrecidos por James Lamb. Como muestra significativa analicemos un caso acerca de un socorrista:

“Supongamos que un socorrista que ha mentido sobre su cualificación no sabe nadar. Asumamos ahora que se ahoga un niño cuya vida era obligación del socorrista salvar. Ciertamente haríamos responsable al socorrista y sin embargo, al no saber nadar, este no podría haber salvado la vida del niño” (Lamb 1993, p. 525). En este ejemplo observamos que un agente realiza una acción que le sitúa en una tesitura que le hace moralmente responsable de una acción posterior en la cual no tiene posibilidades alternativas. En consecuencia, según Lamb, las posibilidades alternativas no son necesarias para la responsabilidad moral.

Los ejemplos estilo Frankfurt consisten en ejemplos de casos planteados primeramente por Frankfurt (1969) y otros más elaborados presentados por él mismo y por otros autores (*e.g.*, Mele y Robb 1998). Frankfurt afirma que lo que él denomina “principio de posibilidades alternativas” (1969, p. 829) es falso. El autor habla de posibilidades alternativas en lo referente a las decisiones morales, sin que por ello haya de existir relación directa de estas con la libertad. Afirma que una persona puede ser perfectamente responsable de sus acciones aunque no hubiera podido actuar de manera distinta a como lo hizo. Con este objetivo, ofrece el ejemplo de un personaje ficticio llamado Jones. Este ejemplo, al igual que todos los del mismo estilo, contempla la presencia de un controlador que consigue que el agente (la persona manipulada) haga lo que él desea que haga, si bien ese mismo controlador no interviene si sabe que el agente se dispone a hacer lo que él quiere que haga. Basándose en esta situación, Frankfurt y sus seguidores afirman que en los casos donde el controlador no interviene porque el agente realizará la acción deseada, dicho agente es moralmente responsable de su acto, aunque de hecho no disponía de posibilidades alternativas. Por lo tanto, el principio de posibilidades alternativas sería falso: las posibilidades alternativas no son necesarias para la responsabilidad moral.

Entre los defensores de los argumentos de Frankfurt se encuentra John Martin Fischer, quien considera, al igual que aquel, que la responsabilidad

moral no necesita de la existencia de posibilidades alternativas, pero en cambio opina que estas son necesarias para la libertad. Fischer es compatibilista en lo referente a la responsabilidad moral, pero incompatibilista en cuanto al libre albedrío se refiere: es un partidario del *semicompatibilismo* (Fischer 2007). Otro semicompatibilista, si bien en el sentido opuesto, es Bruce Waller, quien ha defendido que el libre albedrío es compatible con el determinismo, pero no así la responsabilidad moral (Waller 1990). El semicompatibilismo, pues, desliga la relación directa que con frecuencia se establece entre nuestra libertad y nuestra responsabilidad moral, y afirma que, o bien el libre albedrío es compatible con el determinismo, y no así la responsabilidad moral, o bien viceversa.

Dejando a un lado las posibilidades alternativas, la incompatibilidad entre determinismo y libertad podría ser presentada como conclusión de las premisas (c) y (d), centradas en el control último de nuestras decisiones y/o acciones,<sup>3</sup> y según las cuales: (c) el determinismo impide el control último sobre nuestras decisiones y/o acciones, y (d) el control último es necesario para decidir y/o actuar libremente. Si aceptamos (c) y (d) obtendremos como resultado que el determinismo impide las decisiones y/o acciones libres. Las explicaciones compatibilistas centradas en las posibilidades alternativas se basaban en la negación de (a) o (b). Pero en el caso que nos ocupa tanto compatibilistas como incompatibilistas están de acuerdo en la validez de la premisa (c): hay un consenso general en afirmar que determinismo y control último son incompatibles, ya que como agente sólo puedo tener un control último sobre mis decisiones y/o acciones si estas poseen un origen último causalmente indeterminado o si la causa soy yo mismo.<sup>4</sup> No hay consenso, en cambio, para (d). Las visiones compatibilistas centradas en el control último se

---

<sup>3</sup> Immanuel Kant llamaba “autonomía” a la condición de control último, al igual que Susan Wolf ya en el siglo XX. Otros autores han preferido el término “autodeterminación”.

<sup>4</sup> Llegados a este punto, cabe puntualizar que las teorías de identidad personal narrativa, que otorgan un importante papel al ambiente social en la construcción de la identidad, no conciben el “yo” como agente causal al modo tradicional.

basan en su negación, a saber, rechazan que el control último sea necesario para decidir y/o actuar libremente, si bien afirman que los seres humanos gozamos de un tipo de control *no último* sobre nuestras decisiones y/o acciones. Las principales explicaciones de este tipo proceden de los compatibilistas clásicos y otros como Harry Frankfurt, Gary Watson, Susan Wolf o Daniel Dennett.

Los compatibilistas clásicos, tal y como se ha explicado, conciben la libertad como una capacidad para hacer lo que queremos, deseamos o escogemos hacer, sin que exista nada que nos lo impida. Ya hemos visto qué repercusiones tiene esto sobre la premisa (a). Las repercusiones sobre (d) consisten en que, para ellos, un acto deja de ser libre sólo si está causado por factores externos, y no por los propios deseos y decisiones, que son de carácter interno al agente. Así, el libre albedrío es compatible con el determinismo porque, para gozar de libertad, no es necesario que mis deseos y decisiones sean indeterminados desde el punto de vista causal, sino que simplemente es necesario tener deseos que me hagan capaz de decidir y/o actuar. La libertad no es incompatible con la determinación causal, sino con las coerciones o las constricciones externas al agente. De hecho, los compatibilistas clásicos van más lejos que otros compatibilistas al afirmar que el determinismo es no sólo compatible con, sino también necesario para, la libertad (postura denominada “determinismo blando”).

Harry Frankfurt también ha hecho aportaciones significativas al problema de la relación entre libertad y control último. Para él, la noción de libertad de los compatibilistas clásicos recoge la libertad de acción, pero no de voluntad. Para solventar este problema, ofrece una explicación alternativa (Frankfurt 1971). Muchos animales parecen tener lo que él llama “deseos de primer orden”, es decir, “deseos de hacer o no hacer una cosa u otra” (*ibid.*, p. 7). Pero los humanos gozamos, además, de “deseos de segundo orden”, que son el resultado de un proceso de autoevaluación reflexiva. Estos deseos de segundo orden no serían otra cosa que deseos de tener o no tener este o aquel

deseo de primer orden. Los deseos de segundo orden *controlan*, pues, a los de primer orden, y nuestros deseos de primer orden *son libres* cuando son coherentes con los de segundo orden, cuando tenemos los deseos que *queremos tener*. Se trata de una teoría compatibilista porque no requiere que los deseos de ninguno de los dos niveles estén indeterminados. El control último es, por tanto, innecesario para la libertad.

Además de la de Frankfurt, ha habido otras críticas de los propios compatibilistas hacia el compatibilismo clásico. Actuar tal y como uno quiere actuar no significa necesariamente actuar libremente, pues existen casos de comportamiento adictivo o fóbico que cumplen con el requisito de actuar de ese modo. En definitiva, el compatibilismo clásico fusiona la acción libre y la acción intencionada. Para evitar este problema, Gary Watson (1975) propone hacer una distinción entre *querer* y *valorar*. El individuo posee un *sistema valorativo* consistente en un conjunto de creencias sobre lo que es bueno o adecuado, y cuyo origen se encuentra en su razón. Asimismo, posee un *sistema motivacional*, con origen en su apetito, y que consiste en un conjunto de deseos y motivos de otra índole. La libertad no nace de la conformidad entre deseos de primer y segundo orden, como propone Frankfurt, sino de la conformidad entre el sistema valorativo y el motivacional, entre razón y deseos. Eso explicaría, por ejemplo, por qué las acciones compulsivas de un cleptomaniaco no son acciones libres: su razón no está en conformidad con sus deseos. El control último no es, para Watson, necesario para el libre albedrío porque el agente posee un sistema autónomo de acción que es compatible con el determinismo, pues su razón y sus deseos bien podrían estar determinados sin que ello supusiera un peligro para su libertad.

Susan Wolf (1990) ofrece una explicación que, al igual que la de Watson, apela a los valores del agente, si bien destaca por su alto contenido normativo. Para esta autora, actuar de modo libre requiere la capacidad de distinguir lo verdadero de lo falso y lo bueno de lo malo, que posibilita que el agente pueda hacer *lo correcto* gracias a las *razones correctas*. Así, los humanos

actuamos de modo libre cuando actuamos correctamente debido a razones correctas, pero no así cuando actuamos erróneamente, cuando hacemos lo correcto a causa de razones erróneas o incluso cuando nos equivocamos a la hora de hacer lo correcto debido a razones correctas. Gozamos de un tipo de control que no necesita de la indeterminación para funcionar.

También Daniel Dennett hace una defensa del compatibilismo centrándose en el control último como aspecto clave. Para él, debemos hacer una distinción entre la manipulación del agente por parte de agentes externos y la determinación causada por las fuerzas naturales, puesto que estas últimas no nos manipulan ni controlan aunque determinen nuestras acciones (Dennett 1984). En un trabajo junto a Christopher Taylor (Taylor y Dennet 2002), emplea analogías entre los seres humanos y los ordenadores más sofisticados con la intención de demostrar que la creatividad, la reflexividad y la flexibilidad que los humanos mostramos cuando ejercemos nuestra libertad también son mostrados por estas máquinas. Podemos tener control sobre nuestras acciones y decisiones aunque estén determinadas. Nuestro comportamiento, como el de los ordenadores, puede estar determinado sin que ello suponga una amenaza para nuestro libre albedrío.

Algunos autores han apostado por posturas compatibilistas que no están centradas en las posibilidades alternativas ni en el control último. Por lo tanto, las premisas *(a)*, *(b)*, *(c)* y *(d)* no pueden dar cuenta de posturas de este tipo. Tal es el caso de la de Peter Strawson (1962), quien relaciona la responsabilidad moral de una persona con la adopción de una serie de actitudes reactivas hacia ella misma y las demás: resentimiento, gratitud, indignación, admiración, etc. Sin importar que estas actitudes estén determinadas por nuestras características físicas, biológicas y/o psicológicas, ser moralmente responsable consiste en estar sujeto de forma apropiada a estas actitudes, ser un adecuado candidato a mostrarlas y/o recibirlas en un contexto social en el cual dichas actitudes juegan un importante rol. Strawson piensa que no es necesario adoptar el incompatibilismo para estar a favor de la

responsabilidad moral porque, en el caso de que el determinismo fuera cierto, los seres humanos no dejaríamos de vernos a nosotros mismos y a los demás en un contexto de actitudes reactivas. El determinismo no es relevante para la responsabilidad moral. A esta postura, que posee un alto componente normativo, le han seguido otras en la misma línea, como la de R. Jay Wallace, quien sostiene (1994, p. 1): “Ser un agente moralmente responsable, creo, no es realmente un problema de poseer libertad de voluntad. Más bien implica principalmente una forma de competencia normativa: la habilidad para comprender y aplicar razones morales, y para gobernar la propia conducta a la luz de dichas razones”.

### 1.3. Incompatibilismo

Como contraposición a las explicaciones compatibilistas, ha habido autores que han optado por la opción incompatibilista: el libre albedrío y/o la responsabilidad moral son incompatibles con el determinismo. Existen diferentes versiones incompatibilistas, que ahora expondremos.

El libertarismo es la tesis según la cual los seres humanos tenemos libre albedrío y además este es incompatible con el determinismo.<sup>5</sup> Los libertaristas creen, en consecuencia, que el determinismo no es una ley universal. Autores destacados de esta corriente son Roderick Chisholm, Randolph Clarke, Carl Ginet, Robert Kane, Hugh McCann, Robert Nozick y Timothy O'Connor, entre muchos otros. Estos autores basan sus teorías en la aceptación de las premisas (a), (b), (c) y (d), es decir, piensan que el determinismo excluye las posibilidades alternativas y el control último y que ambas condiciones son necesarias para la libertad.

---

<sup>5</sup> No debemos confundir el libertarismo en el contexto de los debates sobre el libre albedrío con el libertarismo como corriente política.

Además de atender al determinismo y a la incompatibilidad de este con la libertad, los libertaristas deben encontrar la respuesta adecuada a una cuestión fundamental para sus argumentos: ¿es inteligible un libre albedrío indeterminista? Dicha pregunta es conocida como “Cuestión de Inteligibilidad” (Kane 1996, p. 13), y su importancia no debe ser menospreciada. El indeterminismo puede ser una amenaza tan peligrosa para la libertad como pueda serlo el determinismo. En efecto, el indeterminismo supone aceptar que en un instante de tiempo  $t$  existe más de un resultado posible como consecuencia de los eventos pasados. Dicho de otra forma: un mismo estado de las cosas en un momento pasado, bajo el reinado de las mismas leyes naturales, puede desembocar en diferentes estados futuros. Esto puede conducirnos a pensar que, en un suceso indeterminista, todo se reduce al puro azar. De modo análogo al lanzamiento de un dado, en el que su autor parece carecer de control sobre el resultado, el azar parece excluir la autoría de las decisiones y/o acciones: el control último por parte del agente no existiría y el libre albedrío no sería posible. Los libertaristas han de encontrar un tipo de indeterminismo “relevante [...] en el sentido de que *genera* la no-aleatoriedad, o la *consigue*, o la *realza*, o la *incrementa*, o algo similar” (Balaguer 2010, p. 10). Así pues, si el libertarismo quiere ofrecer una teoría satisfactoria debe demostrar, entre otras cosas, que el libre albedrío indeterminista es inteligible.

Existen diferentes versiones del libertarismo. Algunas son dualistas en cuanto a la relación mente-cerebro; otras son materialistas. Unas postulan un tipo de causalidad basado en un agente que es irreductible a su vez a eventos causales; otras postulan una causalidad basada en eventos. Existen versiones “valerianas”,<sup>6</sup> que sitúan la indeterminación relevante para una decisión previamente a esta; y existen versiones “no valerianas”, las cuales sitúan la indeterminación relevante simultáneamente a la decisión.

---

<sup>6</sup> “Valerian” en inglés. Se trata de un término acuñado por Daniel Dennett, en alusión al poeta francés Paul Valéry.

Los “deterministas duros”, llamados así por William James, aceptan las premisas (a), (b), (c) y (d), al igual que los libertaristas, si bien con una finalidad bien distinta: creen que el libre albedrío es incompatible con el determinismo y que los seres humanos no tenemos libre albedrío porque el determinismo es cierto. Figuras destacables de esta corriente son Baruch Spinoza (en el siglo XVII), el Barón de Holbach (en el siglo XVIII) y Clarence Darrow (ya en el siglo XX). Hoy en día los partidarios del determinismo duro no son muy numerosos. La posición sostenida por Ted Honderich es, sin embargo, un incompatibilismo cercano al determinismo duro. Cree que el indeterminismo descrito por la mecánica cuántica no es contradictorio con el determinismo a nivel del comportamiento humano: los efectos cuánticos serían anulados a un nivel macroscópico. Por tanto, los seres humanos no tenemos libertad porque es incompatible con el determinismo y porque este es cierto al menos a escala macroscópica (Honderich 2002).

El “incompatibilismo duro”, tal y como lo ha venido a bautizar Derk Pereboom, es la tesis según la cual los seres humanos no tenemos libre albedrío porque este es incompatible tanto con el determinismo como con el indeterminismo. Según esta definición, será incompatibilista duro quien concluya que determinismo y libertad son incompatibles a partir de la aceptación de (a) y (b), o de (c) y (d), o de las cuatro premisas conjuntamente, sosteniendo a su vez que indeterminismo y libertad son igualmente incompatibles.

Pereboom (2001, 2007) sostiene una versión del incompatibilismo duro basada en la conjunción de dos tesis. La primera de ellas es que nuestras elecciones y acciones son eventos de uno de estos tres tipos: (i) ajeno-deterministas (“alien-deterministic”), que están determinados causalmente por factores que no están bajo el control del agente, (ii) verdaderamente aleatorios, que no son producto de nada ni de nadie, o (iii) parcialmente aleatorios, que son una mezcla de los dos anteriores, es decir, se producen como resultado de factores ajenos al control del agente que contribuyen a producirlos pero que no

los determinan. La segunda tesis es que el tipo de libertad necesaria para la responsabilidad moral no sólo es incompatible con el determinismo, sino con todo tipo de eventos: ajeno-deterministas, verdaderamente aleatorios y parcialmente aleatorios. Juntas, ambas tesis conducen a la conclusión de que no poseemos el libre albedrío requerido para ser moralmente responsables. En otras palabras, el indeterminismo es una amenaza tan seria como pudiera serlo el determinismo, siempre y cuando descartemos la causalidad del agente, que si bien es una posibilidad que no debemos desechar, no resulta creíble si consideramos las mejores teorías físicas de las que disponemos actualmente.

Pereboom niega la premisa (b) y acepta las premisas (c) y (d), pues afirma que las posibilidades alternativas no son cruciales para la libertad y la responsabilidad moral (de hecho, acepta los ejemplos estilo Frankfurt), dejando este papel para el control último, el cual considera necesario para el libre albedrío e incompatible con el determinismo. Sin embargo, considera que el indeterminismo excluye el control último (de hecho, cualquier tipo de control) sobre nuestras decisiones y/o acciones, debido a que el indeterminismo conlleva el azar y no existe control sobre algo que es azaroso. Este argumento es conocido como “argumento de *Mind*”, expresión acuñada por van Inwagen (1983) en alusión a la revista *Mind*, donde se han ofrecido con frecuencia argumentos de este tipo. Cree Pereboom que su postura debe conducirnos hacia una revisión de la manera en que concebimos nuestros juicios morales, nuestras actitudes reactivas, nuestras relaciones interpersonales, etc. Las consecuencias, además, no serían tan destructivas como podría parecer, pues quedarían compensadas por un tratamiento más humano hacia nosotros mismos y hacia los demás.

Existen dos autores con posturas destacables que pueden considerarse cercanas a la del incompatibilismo duro. Se trata de Galen Strawson y Saul Smilansky. Strawson (1986, 1994) afirma que no importa si el determinismo es cierto o no: el tipo de libertad descrito por los libertaristas no es posible. Esto se debe a que, para su funcionamiento, el libertarismo necesita, bien una serie

infinita de regresiones en la cadena causal (unas elecciones libres que preceden a otras y mediante las cuales nos formamos a nosotros mismos), o bien sucesos azarosos sobre los cuales carecemos de control. Strawson también es crítico con los compatibilistas (entre los que se encuentra su padre, Peter Strawson), pues considera que no han logrado explicar convincentemente un libre albedrío en el que esté implicada la responsabilidad moral. Smilansky (2000) propone, por su parte, dos tesis diferenciadas: el “dualismo fundamental” y el “ilusionismo”. El dualismo fundamental postula, por un lado, que el libre albedrío libertarista es imposible, ya que no se ha ofrecido hasta la fecha ninguna explicación convincente de la noción de control último del tipo libertarista. Por otro lado, afirma que no debemos escoger entre el compatibilismo y el determinismo duro, sino que debemos adoptar una explicación intermedia que integre los aciertos de ambas corrientes, ya que ninguna de las dos se sostiene por sí sola de modo convincente. En cuanto a la tesis del ilusionismo, se basa en la creencia de que es necesario, desde los puntos de vista social y moral, mantener la ilusión de que el libre albedrío existe, pues sería nefasta para la humanidad una creencia contraria extendida de modo generalizado. Smilansky no considera que sea necesario introducir dicha ilusión, dado que está ampliamente extendida entre nosotros: o bien no discutimos que tengamos un libre albedrío del tipo libertarista, o bien consideramos que gozamos de la libertad que necesitamos al modo compatibilista. Lo necesario es, por tanto, mantener la ilusión.

#### **1.4. Posturas rupturistas**

Desmarcándose de las posturas habituales, compatibilismo e incompatibilismo, hay autores que han ofrecido teorías alternativas que tratan de romper con la dicotomía entre ambas, no siendo adecuado desgranar dichas teorías a partir de

las premisas (a), (b), (c) y (d). Entre estos autores cabe destacar a Alfred Mele, Richard Double y Manuel Vargas.

Alfred Mele (2002) propone una postura que él llama “autonomismo agnóstico”. En primer lugar opone *autocontrol* y *debilidad de voluntad*. La debilidad de voluntad es, en este sistema, un fallo del autocontrol. En segundo lugar, Mele opina que para que un agente sea autónomo necesita una serie de condiciones añadidas al autocontrol, puesto que este no es suficiente por sí mismo. Pero lo más importante es que aceptar dichas condiciones para la autonomía (y en consecuencia para la libertad) no requiere una decisión entre compatibilismo e incompatibilismo, pues se pueden describir condiciones que ofrecen una explicación convincente tanto de una autonomía compatibilista como de una incompatibilista. Esto lleva a la conclusión, piensa Mele, de que se puede ser agnóstico en lo relativo al debate compatibilismo/incompatibilismo y aún así mantener la convicción de que existen personas autónomas. La propuesta de Mele conduce, pues, a la aceptación de la existencia del libre albedrío humano, sin importar si son ciertos el compatibilismo o el incompatibilismo.

Richard Double (2002) defiende una postura que él denomina “subjetivismo”. Considera que el debate en torno al libre albedrío se debe principalmente a la adopción de diferentes posturas *metaéticas*. La responsabilidad moral se basa en propiedades de las personas que no son objetivas, sino que expresan nuestras actitudes morales particulares (y por lo tanto subjetivas) hacia el comportamiento y los actos ajenos. Y en la medida en que Double opina que por “libre albedrío” entendemos el tipo de libertad que se requiere para poseer responsabilidad moral, y que es algo que una persona podría tener pero también no tener, piensa que atribuir la capacidad de libre albedrío a una persona es un hecho subjetivo basado en dichas actitudes morales.

Manuel Vargas califica su postura como “revisionista”. Considera que, al igual que la humanidad ha ido modificando a lo largo del tiempo su idea

acerca de conceptos como “agua”, “matrimonio” o “mago”, podemos *revisar* nuestros conceptos de libre albedrío y responsabilidad moral, abandonando aquellos aspectos que nos resulten más problemáticos (Vargas 2007, pp. 126-8). Pueden existir diversas teorías revisionistas: una persona es partidaria del revisionismo si propone que lo que deberíamos creer acerca de la libertad y/o de la responsabilidad moral se desmarca de lo que normalmente tendemos a creer (esto último es lo que Vargas conoce como “sentido común”). De esta manera, “uno puede sostener [...] que el sentido común es compatibilista, pero que deberíamos creer en una concepción incompatibilista del libre albedrío”, lo que define como “incompatibilismo revisionista” (*ibid.*, pp. 151-2). De igual manera, se puede creer que el sentido común es incompatibilista acerca del concepto de libertad y/o el de responsabilidad moral, y proponer una revisión compatibilista como explicación adecuada a dichos conceptos. Esta última es, precisamente, la postura de Vargas, quien ofrece su propia versión compatibilista del libre albedrío y de la responsabilidad moral, versión que, a su juicio, debería sustituir a nuestra idea ordinaria, la cual contiene elementos incompatibilistas que deberíamos revisar (Vargas 2007, 2009).



## Capítulo 2. Explorando el nivel “micro”: mecánica cuántica y libre albedrío

«Vosotros no podéis comprender lo que significa  
llevar en el interior dos voces ¡como yo llevo!,  
¡gritando!, ¡gritándome constantemente!  
“¡No lo hagas! ¡Mata! ¡¡No lo hagas!! ¡¡Mata!!”  
Y las voces... ¡siguen enloqueciéndome!  
¡Y yo quiero impedirlo, pero no puedo evitarlo!»

*M, el vampiro de Düsseldorf* (Fritz Lang, 1931).<sup>1</sup>

En el año 1900, el científico alemán Max Planck planteó una hipótesis que cambiaría para siempre el devenir de la física. Para Planck, es imposible emitir ondas, tales como los rayos X o la luz, en cantidades arbitrarias. Sólo es posible hacerlo en paquetes a los que denominó “cuantos”. Esta hipótesis constituiría el inicio de una nueva rama de la física, la mecánica cuántica, que tendría profundas implicaciones en el modo en que los humanos concebimos el universo. Para los objetivos de este trabajo, no obstante, las implicaciones más destacables de la mecánica cuántica guardan relación con el determinismo universal, para el que podría suponer una seria amenaza.

---

<sup>1</sup> El guión de esta película es obra de Thea von Harbou y el propio Fritz Lang. El fragmento transcrito corresponde a la versión doblada al castellano.

## 2.1. Aspectos fundamentales de la mecánica cuántica. Interpretación de Copenhague

En la presente sección expondremos los aspectos esenciales de la mecánica cuántica, y lo haremos siguiendo el enfoque adoptado por la interpretación de Copenhague. Esta interpretación, la más comúnmente aceptada, posee un marcado tono indeterminista.

### 2.1.1. Descubrimiento de los cuantos

Planck (1900) descubrió que la energía de las ondas electromagnéticas no puede dividirse en fracciones, sino que se encuentra en paquetes, a los que denominó “cuantos”, cuyo contenido es múltiplo de una unidad fundamental. Es más, el paquete de energía mínimo que puede presentar la onda, su unidad de energía, es directamente proporcional a su frecuencia (la cual, a su vez, es inversamente proporcional a la longitud de onda). Esta relación puede ser cuantificada mediante un factor llamado “constante de Planck” ( $h$ ), cumpliéndose la igualdad:

$$E = h\nu \text{ ,}$$

donde  $E$  es la unidad de energía mínima de la onda electromagnética,  $h$  es la constante de Planck y  $\nu$  es la frecuencia de onda. El valor de  $h$  es minúsculo (aproximadamente  $6,63 \cdot 10^{-34}$  J·s), de forma que los paquetes de energía tienen normalmente un tamaño pequeño. Los cambios de energía, igualmente, se producen en pequeños saltos, si bien estos son mayores conforme crece la frecuencia de onda (esto es, conforme disminuye la longitud de onda).

El planteamiento de Planck fue clave para las ideas de Niels Bohr (1913a, 1913b, 1913c) acerca de la estructura de los átomos, con los cuantos

como elementos esenciales de esta. Sin embargo, Planck no encontró la forma de explicar por qué la energía se presenta en paquetes, algo que sí consiguió Albert Einstein investigando acerca del efecto fotoeléctrico. Este fenómeno consiste en la emisión de electrones por parte de algunos metales cuando son iluminados con radiación electromagnética. Cuando esta radiación choca con la superficie del metal, le transmite su energía, agitando algunos de los electrones de esta capa que están débilmente unidos a los átomos. Esta agitación provoca que los electrones puedan ser despedidos de la superficie metálica.

El efecto fotoeléctrico posee ciertas características especiales. La intensidad de la radiación no influye de ningún modo en la velocidad de emisión de los electrones. Un aumento de la intensidad provoca que exista un mayor número de electrones emitidos, pero no consigue que aumente la velocidad a la que salen despedidos de la superficie del metal. Esta sí crece, por el contrario, cuando la frecuencia de la luz (radiación electromagnética) es mayor, es decir, cuando es menor su longitud de onda. También sucede lo contrario: un descenso en la frecuencia conlleva una disminución en la velocidad de emisión. Incluso llega un momento en que, cuando la frecuencia de la luz es lo suficientemente baja, los electrones dejan de ser emitidos, sin importar en ningún momento cuál sea la intensidad de la luz.

La aportación que Einstein realizó al estudio del efecto fotoeléctrico (Einstein 1905) parte de la idea de integrar en la estructura de la luz la explicación de Planck acerca de los cuantos de energía de las ondas. Si dividimos la onda luminosa en paquetes, es decir, partículas de luz (a las que posteriormente se denominaría “fotones”), podemos encontrar una explicación al efecto fotoeléctrico. Cada fotón posee una energía que es proporcional a la frecuencia de onda, y esta ha de alcanzar un valor mínimo para que los electrones comiencen a ser emitidos desde la superficie metálica. En otras palabras, los fotones han de llevar asociada una energía mínima para que se produzca el efecto fotoeléctrico. Por otro lado, si aumentamos los electrones que contiene un rayo de luz, crece la intensidad de este, provocando que un

número mayor de electrones sean agitados y emitidos, si bien su velocidad de emisión permanecerá constante mientras no varíe la frecuencia de la onda luminosa (y con ella la energía de los fotones). Einstein, de esta manera, dejó patente la eficacia de la explicación cuántica de Planck a la hora de describir la radiación electromagnética.

### *2.1.2. Dualidad onda-partícula*

Isaac Newton afirmó que la luz está compuesta por partículas, pero su colega Thomas Young demostró, mediante el célebre experimento de la doble rendija (figura 2.1.), que el comportamiento de la luz obedece al de las ondas. Explicaremos a continuación la formulación moderna de este experimento, elaborada a partir de su formulación original. En una cámara oscura colocamos una fuente luminosa que dirigirá un haz de luz hacia una placa fotográfica situada en el lado opuesto. Entre ambas ubicamos una pared con dos rendijas verticales y paralelas entre sí. Si tapamos una de las dos rendijas, en la placa fotográfica quedará reflejada una única franja brillante como resultado del paso de la luz por la otra rendija, que permanece abierta. Pero si ambas rendijas están abiertas, la imagen reflejada en la placa no consistirá en dos franjas brillantes paralelas entre sí, como cabría esperar si la luz siguiera un comportamiento corpuscular. En su lugar, lo que se observa es un característico diagrama de interferencias, con alternancia de franjas oscuras y brillantes. Al interactuar la luz de ambas rendijas hay una interferencia de las ondas de modo que las crestas de algunos conjuntos de ondas se superponen con los valles de otros conjuntos, anulándose entre sí. Este tipo de interferencia produce las franjas oscuras. Sin embargo, otros conjuntos de ondas interfieren de modo que crestas y valles coinciden, aumentando la intensidad de la onda luminosa y produciendo las franjas brillantes.

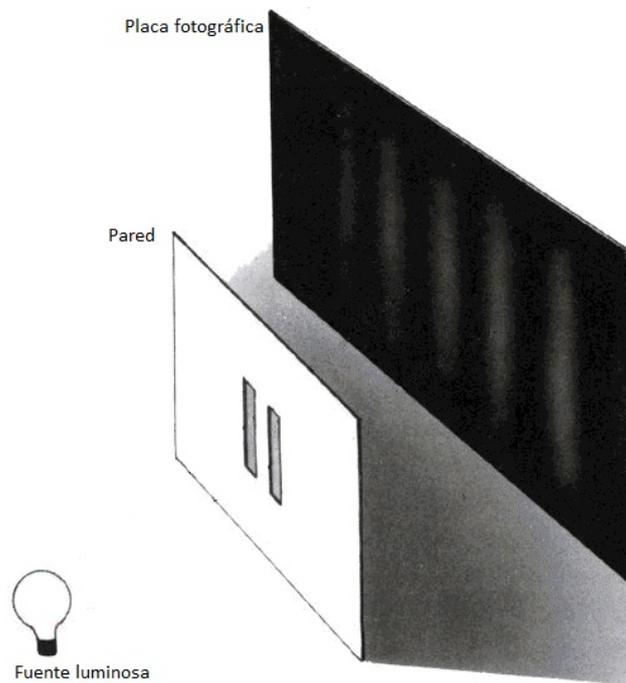


Figura 2.1. Experimento de la doble rendija: la interferencia entre conjuntos de ondas produce un característico patrón con alternancia de bandas brillantes y oscuras. Modificado de Hawking (1988).

El experimento de la doble rendija deja patente la naturaleza ondulatoria de la luz, pero las investigaciones de Einstein acerca del efecto fotoeléctrico demuestran igualmente su carácter corpuscular. Parece que los fotones, pese a ser partículas, se comportan de modo parecido a como lo hacen las ondas. De hecho, tal y como hemos podido ver, la energía de los fotones depende de la frecuencia, una característica que presentan las ondas. Basándose en todo esto, la mecánica cuántica proclama que la radiación electromagnética posee una dualidad onda-corpúsculo (dualidad onda-partícula). Según esta, las partículas pueden presentar un comportamiento ondulatorio y las ondas, a su vez, pueden comportarse tal y como lo hacen las partículas. Es más, Louis de Broglie (1925) propuso que la dualidad onda-partícula es una característica no sólo de la luz, sino de toda la materia. A la vista de nuestro conocimiento empírico actual, puede decirse que la propuesta era acertada; sin ir más lejos,

los electrones manifiestan un patrón de interferencias similar al que presenta la luz.

### *2.1.3. No-localidad*

Cuando dos sistemas cuánticos interactúan, permanecen de algún modo correlacionados de manera que el estado de uno de los sistemas influye en el estado que poseerá el otro, y viceversa, sin importar que ambos se encuentren lejanos en el espacio-tiempo. Esta misteriosa correlación fue rechazada por Einstein, Podolski y Rosen (1935), quienes formularon la llamada “paradoja EPR” con el fin de rebatir la no-localidad cuántica. Décadas más tarde, no obstante, Aspect, Dalibard y Roger (1982) demostraron la no-localidad mediante un experimento de polarización <sup>2</sup> con fotones. Observaron que si emitimos dos fotones simultáneamente y medimos la polarización de uno de ellos, existe una correlación con la polarización del otro de forma que el sistema compuesto por ambos es descrito mediante un único estado cuántico. Las polarizaciones de ambos fotones se influyen mutuamente aunque estén separados en el espacio. Pero, ¿qué es un estado cuántico? Lo veremos en el próximo apartado.

### *2.1.4. Principio de incertidumbre*

Si bien los cuantos, la dualidad onda-partícula y la no-localidad son elementos fundamentales de la mecánica cuántica, el punto de partida para comprender la relación entre esta y el determinismo reside en un descubrimiento efectuado

---

<sup>2</sup> Este fenómeno consiste en la oscilación de las ondas electromagnéticas en un solo plano (el plano de polarización), en contraste con la oscilación en todas las direcciones con la que se propaga normalmente una onda.

por Werner Heisenberg (1927). Se trata del principio de incertidumbre, según el cual no podemos extraer simultáneamente, de modo preciso, el valor de dos observables de una partícula (tales como su posición y su momento lineal). La formulación matemática del principio, que adaptamos aquí a partir de la propuesta por Kennard (1927), es la siguiente:

$$\Delta r \Delta p \geq h/4\pi \quad ,$$

donde  $h$  es la constante de Planck, mientras que  $\Delta r$  y  $\Delta p$  corresponden a la desviaciones típicas (incertidumbres) en la posición y en el momento lineal de la partícula, respectivamente. Si queremos averiguar la posición y el momento lineal (o cantidad de movimiento) actuales de una partícula, con el objetivo de predecir su posición y su cantidad de movimiento futuras, debemos iluminar dicha partícula. Seremos capaces de conocer la posición gracias a que algunas ondas electromagnéticas rebotan en ella y llegan a nuestra retina. Sin embargo, en esta determinación de la posición jugamos con un margen de error que es exactamente la distancia entre dos crestas consecutivas de la onda electromagnética, es decir, la longitud de onda. Por esta razón, para aumentar la precisión en el estudio de la posición será conveniente aplicar radiación con longitud de onda muy corta, o lo que es lo mismo, con muy alta frecuencia. Además, la mínima cantidad de luz que podemos aplicar durante la observación corresponde a un fotón (un cuanto de luz). El problema reside en que los fotones con corta longitud de onda albergan una gran cantidad de energía, de forma que al rebotar con la partícula en cuestión provocan inevitablemente una alteración en la cantidad de movimiento de esta. Si, por el contrario, utilizamos luz de alta longitud de onda, cuyos fotones poseen una menor cantidad de energía, la alteración sobre el momento lineal de la partícula observada será pequeña y podremos determinar con precisión dicho momento lineal. En este caso, el problema surgirá a la hora de averiguar la posición, puesto que el margen de error será mayor, dado que es igual a la longitud de

onda, la cual será grande. En definitiva, cuanto mayor sea la precisión (menor la incertidumbre) con que hallemos la posición de una partícula, menor será la precisión (mayor la incertidumbre) con la que determinaremos su cantidad de movimiento; al revés, la relación funciona de la misma manera. Matemáticamente:

$$\Delta p \geq h / 4 \pi \Delta r ,$$
$$\Delta r \geq h / 4 \pi \Delta p .$$

Por si fuera poco, la incertidumbre se presenta, según Heisenberg, independientemente del tipo de partícula, del método utilizado y de las propiedades estudiadas. Por ejemplo, hay una relación inversa entre la precisión en la medida de la cantidad de energía de una partícula y el tiempo invertido en realizar dicha medición: no es posible conocer la energía de una partícula en un instante preciso. A raíz de todas estas observaciones, el estudio de las partículas en mecánica cuántica no se realiza a partir de la descripción separada de sus observables (como por ejemplo su posición y su momento lineal), sino de una combinación de ellas denominada “estado cuántico”.

#### 2.1.5. Ondas y probabilidad

Max Born (1926) realizó una aportación a la teoría cuántica basada en estudiar las ondas electrónicas en términos probabilísticos: existen lugares de la onda donde es más probable encontrar el electrón que en otros. Esta explicación, ampliada a toda la materia, supone que esta, por su naturaleza ondulatoria, puede ser descrita mediante probabilidades. En los niveles macroscópicos esta descripción no tiene relevancia, puesto que la cualidad ondulatoria de la materia no se puede percibir, pero si nos centráramos en los niveles microscópicos nos encontraríamos con que la mejor manera de estudiar la

posición de una partícula es asignarle probabilidades particulares de encontrarse en distintos lugares. No en vano, *la repetición de un experimento bajo condiciones idénticas no garantiza idénticos resultados*. Imaginemos que hacemos rebotar una onda electrónica contra una pared. Los electrones, tras encontrarse con el obstáculo, podrían encontrarse en lugares diversos para los cuales disponemos de probabilidades específicas. Si a continuación decidimos repetir el experimento en condiciones exactamente iguales a las del primer intento, posiblemente tendremos un resultado distinto en lo que se refiere a la posición de los electrones. Como observadores del experimento somos absolutamente incapaces de predecir la posición concreta en que se encontrará un electrón. En su defecto, no tenemos otro remedio que asignar una probabilidad de que este, en efecto, se halle allí. Las predicciones pueden efectuarse gracias a la ecuación de Schrödinger, surgida a raíz de la investigación de este físico sobre la mecánica ondulatoria de las partículas (véanse Schrödinger 1926a, 1926b, 1926c, 1926d). Puede expresarse de la siguiente forma:

$$\hat{H}\Psi = i (h/2\pi) (d\Psi/dt) ,$$

donde  $\hat{H}$  es el operador hamiltoniano (referente a la energía total),  $i$  es la unidad imaginaria ( $i = \sqrt{-1}$ ) y  $h$  es la constante de Planck. La ecuación de Schrödinger muestra cómo evoluciona la “función de onda” de un sistema cuántico ( $\Psi$ ), término acuñado para las ondas probabilísticas ya mencionadas. Ampliamente utilizada, esta ecuación ha dado lugar a predicciones muy eficaces en la investigación científica y tecnológica. Permite predecir la probabilidad de alcanzar diferentes futuros posibles (aunque, eso sí, no hay modo de saber cuál de ellos sucederá). Debemos aclarar, no obstante, que la probabilidad de que una partícula se halle en un lugar concreto en un momento determinado no viene dada por la función de onda, sino por el cuadrado de su valor absoluto,  $|\Psi|^2$ . También es importante tener presente que la ecuación de Schrödinger no constituye la única opción disponible a la hora de efectuar predicciones

cuánticas. Existen otras vías, como la suma de historias posibles de Feynman, sobre la cual hablaremos en el apartado 2.2.1.

¿Cómo debe ser interpretada la probabilidad cuántica? A este respecto, Heisenberg (1958, p. 42) sostiene:

La probabilidad en matemáticas o en mecánica estadística significa una declaración sobre nuestro grado de conocimiento de la situación real. Al lanzar el dado no conocemos los sutiles detalles del movimiento de nuestras manos que determinan la caída del dado y, por tanto, decimos que la probabilidad de sacar un número en particular es exactamente de uno entre seis. La onda de probabilidad [...], sin embargo, significaba más que eso; significaba una tendencia hacia algo. Era *una versión cuantitativa del viejo concepto de 'potencia' en la filosofía aristotélica* [cursiva añadida]. Introducía algo situado en el medio entre la idea de un evento y el evento real, un extraño tipo de realidad física justo en el medio entre la posibilidad y la realidad.<sup>3</sup>

#### 2.1.6. Problema de la medida

Mientras que la física clásica supone que efectuar una medida saca a la luz propiedades intrínsecas del sistema, presentes en él antes de medirlo, la interpretación de Copenhague postula lo contrario: no tenemos razones para creer que las propiedades estaban presentes antes de que midiéramos el sistema. Por el contrario, este se encontrará en una superposición de estados que sólo se resolverá al medir, revelando un estado concreto. Este hecho provoca la aparición de lo que llamamos “problema de la medida” en mecánica cuántica. La razón de que sea un problema es que, en ausencia de medición, se considera que la función de onda que describe al sistema evoluciona de modo

---

<sup>3</sup> En los últimos tiempos, un autor que ha venido realizando una lectura aristotélica en la línea de la que hemos citado es Christian de Ronde (e.g., de Ronde 2012).

determinista, pero no puede explicar porqué la medida provoca una reducción desde una superposición de estados a uno solo en concreto.<sup>4</sup>

Una útil ilustración de este problema es una paradoja conocida como “el gato de Schrödinger”. Se trata de un experimento imaginario concebido por el propio Schrödinger (1936). Pensemos en una caja en cuyo interior se aloja un gato. En ella encontramos también un recipiente que contiene una sustancia altamente venenosa, un detector de desintegración radiactiva y un martillo que interconecta ambos de manera tal que, cuando el detector refleja una desintegración radiactiva, el martillo rompe el recipiente y el veneno es liberado, matando al animal. Si no hay desintegración, no se libera veneno alguno y el gato continúa con vida. Cuando el experimento se va a llevar a cabo, se introduce en la caja una sustancia radiactiva que tiene un 50 % de posibilidades de que uno de sus átomos se desintegre en el transcurso de una hora. Cuando la hora ha pasado, habrá la mitad de posibilidades de que el gato esté vivo y la mitad de que esté muerto, pero no podremos saber el desenlace si no abrimos la caja. Pues bien, según las leyes mecánico-cuánticas, la función de onda que describiría el estado del gato previamente a la apertura de la caja reflejaría la superposición de los estados “vivo” y “muerto”: el gato no estaría concretamente vivo ni muerto, sino en los dos estados al mismo tiempo. Sólo podremos saber si el gato vive o ha muerto abriendo la caja, y al hacerlo modificaremos el sistema rompiendo la *superposición* de los dos estados, observando que, o bien el gato sigue con vida, o bien ha muerto. En ese momento de apertura de la caja, y sólo en ese momento, el gato estará vivo o muerto. Hasta ese instante estará vivo y muerto *simultáneamente*.<sup>5</sup>

---

<sup>4</sup> En la célebrima Conferencia Solvay de 1927, Paul Dirac defendió que dicha reducción corresponde a una elección tomada por el mundo natural (consúltese Bacciagaluppi y Valentini 2009, p. 166).

<sup>5</sup> Según de Ronde (2012, p. 156),

[un] indicador que expresa la distancia respecto de las superposiciones clásicas se encuentra dado por la constitución de la superposición cuántica a partir de *términos contradictorios*, por

El problema teórico suscitado por la paradoja recién expuesta y el problema de la medición reflejan dejan patente la dificultad de aceptar, sin más, una única interpretación de los fenómenos cuánticos (la de Copenhague en esta ocasión).

### 2.1.7. Desafío al determinismo

En la introducción de la presente tesis ofrecíamos tres definiciones del concepto “determinismo”. La primera era la aportada por Fischer *et al.* (2007, p. 2): “podemos tratar el determinismo como la tesis según la cual en cualquier instante [...] el universo posee exactamente un solo futuro posible físicamente”. Otras dos definiciones tenían la autoría de Peter van Inwagen. Una de ellas (1983, p. 65) era la conjunción de dos tesis: (1) “Para cada instante de tiempo, existe una proposición que expresa el estado del mundo en ese instante”, y (2) “Si  $p$  y  $q$  son proposiciones cualesquiera que expresan el estado del mundo en algunos instantes, entonces la conjunción de  $p$  junto con las leyes de la

---

ejemplo, “gato vivo” y “gato muerto”, “*spin* para arriba” y “*spin* para abajo”. De este modo, si tomamos en cuenta la superposición cuántica más allá de su expresión formal, se nos presenta entonces como fundamental la consideración de la *contradicción* tanto en términos formales como conceptuales o interpretativos. ¿Cómo puede ser pensada la contradicción? Hoy contamos con las lógicas *paraconsistentes*, que ciertamente podrían ser consideradas con el objetivo de avanzar en el camino de un desarrollo de la estructura formal de la teoría.

La lógica paraconsistente se caracteriza por su permisividad hacia las contradicciones, y precisamente de esta índole es el enfoque que el propio de Ronde y Newton da Costa han adoptado en relación a las superposiciones cuánticas (da Costa y de Ronde 2013). Sostienen los autores:

Nuestra propuesta se centra en la idea de que valdría la pena desarrollar una nueva interpretación de las superposiciones cuánticas que tenga en cuenta la contradicción “justo desde el principio”. [Planteamos] una aproximación paraconsistente a las superposiciones cuánticas que muestra la posibilidad de tener en cuenta las contradicciones también desde una perspectiva formal. Sin embargo, debería quedar claro que no asumimos que la lógica paraconsistente sea la “lógica auténtica” que debería reemplazar a la lógica clásica [...]. Desde nuestra perspectiva sostenemos que los físicos deberían reconocer la posibilidad de utilizar nuevas formas de lógica—como la lógica paraconsistente—que podrían ayudarnos a comprender rasgos de diferentes dominios de la realidad [...]. No creemos que haya una “lógica auténtica”, sino más bien que sistemas lógicos distintos pueden ser útiles para desarrollar y comprender aspectos complementarios de la realidad (*ibid.*, pp. 855-6).

naturaleza implica  $q$ ". La otra definición, en la misma obra y página, decía: "El determinismo es, intuitivamente, la tesis según la cual, dados el pasado y las leyes de la naturaleza, existe sólo un futuro posible".

A la vista de las definiciones recién expuestas, resulta fácil visualizar el impacto que el principio de incertidumbre y el problema de la medida poseen sobre el determinismo. ¿Cómo podemos confiar en las leyes de la naturaleza como actor fundamental para alcanzar un solo futuro posible en cada instante de tiempo si existen leyes físicas cuya esencia es la impredecibilidad? Una respuesta posible sería que no tenemos por qué aceptar la impredecibilidad cuántica, pues la mecánica cuántica podría ser una teoría incorrecta o, como creía Einstein, incompleta. De hecho, el gran caballo de batalla de la física actual es la búsqueda de una teoría que integre eficazmente la mecánica cuántica y la relatividad, pues ambas son incongruentes entre sí,<sup>6</sup> lo que puede ser un indicio de que algo va mal en, al menos, una de ellas. Sin embargo, la mecánica cuántica ha constituido un pilar para el desarrollo científico y tecnológico hasta nuestros días,<sup>7</sup> motivo por el cual parece razonable tomarla muy en serio. Otra posible respuesta sería que no tenemos por qué pensar que la impredecibilidad refleje un aspecto esencial del universo, sino más bien nuestra incapacidad tecnológica y/o intelectual para estudiar los niveles más minúsculos del universo. En otras palabras, podríamos pensar que el indeterminismo cuántico no se presenta a nivel *ontológico*, sino sólo a nivel *epistemológico*. Esta posibilidad de elección entre dos interpretaciones distintas del indeterminismo cuántico ha sido destacada por Bishop (2002, pp. 117-18). La causalidad presente en la mecánica cuántica, dice Bishop, es una causalidad de tipo probabilista, y las probabilidades en las que se basa pueden tener dos naturalezas: epistemológica y ontológica.

---

<sup>6</sup> Una amena y accesible introducción a este problema puede encontrarse en Greene (1999).

<sup>7</sup> El láser, los aceleradores de partículas, los superconductores de alta temperatura, la televisión en HD, las comunicaciones vía satélite... son sólo algunos de los avances tecnológicos a los que la mecánica cuántica ha contribuido decisivamente.

Si las probabilidades poseen naturaleza epistemológica, puede ser por dos razones: (1) la presencia de variables ocultas (mecanismos desconocidos) en nuestras observaciones, y (2) la existencia de una interacción con el entorno que no es tomada en cuenta. Entre los partidarios de la opción (1) podemos destacar a Bohm (1952a, 1952b; Bohm y Hiley 1993) o a Bub (1997), mientras que la opción (2) ha sido defendida por Zurek (1981, 1982, 1986, 1991). En cualquiera de los dos casos las ondas de probabilidad surgen de nuestra falta de conocimiento y no de que la indeterminación sea una característica esencial de los sucesos cuánticos. Esto abre la posibilidad de que la mecánica cuántica pueda ser de naturaleza determinista.

Si las probabilidades son de naturaleza ontológica es porque no existe variable, mecanismo o factor alguno que influya en la impredecibilidad de las propiedades de una partícula en un instante determinado. Así lo creen, entre otros, von Neumann (1955), Gisin (1984, 1989) o Diósi (1988). En este caso, las ondas de probabilidad se presentan, ni más ni menos, porque el indeterminismo es una propiedad esencial de la mecánica cuántica, destruyendo la posibilidad del determinismo universal.

## **2.2. Interpretaciones alternativas y determinismo**

Si bien hasta este momento hemos considerado los fundamentos básicos de la mecánica cuántica a través del prisma de su interpretación más ampliamente aceptada (la de Copenhague), existen muchas otras de entre las cuales expondremos tres que parecen abrir la puerta al determinismo cuántico: la

suma de historias posibles de Feynman, la interpretación de los mundos múltiples y la interpretación de Bohm.<sup>8</sup>

### 2.2.1. *La suma de historias posibles y los mundos múltiples*

Richard Feynman (1948) propuso un método denominado “suma de historias posibles”, en el cual se considera que una partícula, en su viaje entre dos puntos cualesquiera, no toma una única trayectoria. En lugar de ello, recorre *todos* los caminos posibles entre ambos puntos. Si pensamos en el experimento de la doble rendija (figura 2.1.), un fotón que parta desde la fuente de luz hacia la placa fotográfica recorrerá una gigantesca cantidad de trayectorias hasta llegar a esta. Hará una línea recta entre fuente luminosa y placa, atravesando la rendija izquierda. Se dirigirá hacia la rendija derecha para atravesarla pero de repente se desviará hacia la izquierda para pasar por ella. Hará un bucle y pasará por ambas rendijas. Llegará hasta la otra punta del país para volver después y atravesar la rendija derecha en dirección a la placa. Y así podríamos continuar indefinidamente. La probabilidad de que el fotón o cualquier otra partícula haga el recorrido entre dos puntos se obtiene sumando *todas las trayectorias posibles* entre ambos. Sin embargo, cuando tratamos con objetos macroscópicos, sucede que las trayectorias se anulan mutuamente exceptuando una sola, que es la que podemos percibir.

---

<sup>8</sup> Otra interpretación considerada habitualmente como determinista es la interpretación de las muchas mentes, que es omitida aquí por estar estrechamente relacionada con la suma de historias posibles y los mundos múltiples. También es digna de mención la interpretación Madhyamika, que guarda similitud con la de Bohm pero se ciñe a lo fenoménico. Además, existen alternativas no deterministas (la mayoría indeterministas y otras agnósticas): la interpretación de von Neumann, las historias consistentes, la interpretación transaccional, las teorías de colapso objetivo, etc. No obstante, el contraste entre la interpretación indeterminista más ampliamente aceptada (la de Copenhague) y las principales alternativas deterministas nos parece suficiente para poder analizar la Cuestión Determinista (apartado 2.6.4.).

La propuesta de Feynman guarda relación con otra interpretación de la mecánica cuántica conocida como “interpretación de los mundos múltiples”, defendida en primer lugar por Everett (1957) y más tarde por otros como Gribbin (1985), Barrow y Tipler (1988), Lockwood (1989) o Deutsch (1998). En esta interpretación se afirma que, al efectuar una medición, el observador se relaciona con el sistema medido de manera que no tiene lugar realmente una reducción desde una superposición de estados a uno solo. Como indica Hodgson (2002, pp. 95-6), si seguimos la interpretación de los mundos múltiples el gato de Schrödinger sería observado tanto vivo como muerto. Las dos observaciones tendrían lugar, si bien una no sabría acerca de la otra.

### 2.2.2. *La interpretación de Bohm*

En la interpretación de David Bohm (1952a, 1952b; Bohm y Hiley 1993) se sugiere la posibilidad de que existan variables ocultas influyendo sobre las probabilidades cuánticas y se niega la influencia del observador sobre los resultados. Las partículas poseen posiciones definidas, sin importar si se realiza una medición sobre ellas o no. Además, existe un campo llamado “campo cuántico”, que afecta al movimiento de las partículas pero que es distinto de otros campos como el gravitatorio o el electromagnético. Esto se debe a que su fuerza no influye sobre su efecto (aunque sí lo hace su forma), es decir, al aumentar la distancia no disminuye el efecto del campo cuántico. Esta propiedad hace impredecibles las posiciones de las partículas, de manera que las probabilidades de la interpretación de Copenhague resultan eficaces a la hora de estudiarlas. El campo cuántico, además, provoca que dos partículas puedan estar correlacionadas pese a estar mutuamente distantes, dando cuenta así del fenómeno de la no-localidad. Ahora bien, no existe evidencia experimental a favor de la existencia de dicho campo. En cualquier caso, la interpretación de Bohm puede ser considerada como determinista porque

supone que la conjunción de los eventos pasados y las leyes naturales conduce a un único futuro posible en cada instante de tiempo.

### **2.3. ¿Existe una correlación entre el indeterminismo cuántico y el libre albedrío?**

A la vista de la posibilidad de que la mecánica cuántica sea indeterminista desde el punto de vista ontológico, cabe preguntarse si los fenómenos cuánticos, así interpretados, pueden constituir un correlato físico del libre albedrío concebido al modo libertarista. Se trata de un debate controvertido, en el que han estado involucrados científicos y filósofos de diversa índole.

#### *2.3.1. Posturas a favor*

Pascual Jordan (1932, 1934, 1938) fue un pionero en la corriente a favor de la correlación entre mecánica cuántica y libertad. Para Jordan, las funciones vitales celulares están dominadas por los fenómenos cuánticos. Si bien se trata de una propuesta insostenible dado nuestro estado actual de conocimiento, constituyó una base para la investigación posterior.

El físico Roger Penrose ha sostenido que la mecánica cuántica podría hacer posible el libre albedrío. Considera (Penrose 1989, 1994) que la consciencia de los seres humanos es en gran parte no-algorítmica, no utilizando en muchas ocasiones mecanismos normativos ni algoritmos para su funcionamiento, y por tanto no puede ser reproducida computacionalmente. Piensa Penrose que las máquinas, que disponen de un número finito de enunciados y reglas, no podrían funcionar exactamente igual que una mente

humana.<sup>9</sup> En cuanto a la consciencia, sería un requisito para la capacidad de comprensión, y esta a su vez para la racionalidad. Pero, ¿qué tipo de fenómenos físicos podrían explicar una consciencia no-algorítmica? La respuesta se dirige hacia el proceso de reducción desde una superposición de estados cuánticos a uno solo, proceso que, para Penrose, no está bien explicado por la interpretación de Copenhague. Dicha interpretación, recordemos, trata de describir la reducción mediante una explicación basada en probabilidades. El autor opina, sin embargo, que este fenómeno podría ser descrito mediante una explicación no-algorítmica, tal y como la necesaria para explicar la racionalidad y la consciencia de la que esta depende. Sugiere, junto con Stuart Hameroff, la existencia de una “reducción objetiva orquestada” (“Orch OR”): un conjunto de reducciones objetivas, no-locales y sincronizadas, entre los microtúbulos de numerosas neuronas cerebrales (Hameroff y Penrose 1996). En su teoría,<sup>10</sup>

los estados cuánticos superpuestos surgen en las proteínas de las subunidades de los microtúbulos (‘tubulinas’), permanecen coherentes, y reclutan más tubulinas superpuestas hasta que se alcanza un umbral de masa–tiempo–energía (relacionado con la gravedad cuántica). A esa altura, el auto-colapso, o reducción objetiva (**OR**), sucede abruptamente. Equiparamos la fase de pre-reducción, de superposición coherente (‘computación cuántica’), con los procesos pre-conscientes, y cada **OR** instantánea (y no computable), o auto-colapso, con un evento consciente discreto. Las secuencias de eventos **OR** dan lugar a un ‘flujo’ de consciencia. Las proteínas asociadas a microtúbulos pueden ‘afinar’ las oscilaciones cuánticas de los estados superpuestos coherentes; la **OR** está por consiguiente auto-organizada, u ‘orquestada’ (‘**Orch OR**’). Cada evento **Orch OR** selecciona (de manera no computable) estados de las subunidades de los microtúbulos que regulan funciones sinápticas/neurales utilizando la señalización clásica (*ibid.*, p. 36).

---

<sup>9</sup> Penrose se inspira en una teoría presentada en su día por J.R. Lucas (1961).

<sup>10</sup> Puede leerse una descripción reciente de la teoría *Orch OR* en Hameroff (2012).

Se trata de un fenómeno de naturaleza no-algorítmica que, cree Penrose, podría constituir una base física no sólo para la consciencia sino también para el libre albedrío.

Henry Stapp (1993, 1998, 1999, 2007) cree que la interacción entre el observador y el mundo que lo rodea es fundamental para entender la consciencia y podría ser importante para comprender el libre albedrío. Según Stapp, las propiedades cuánticas de las partículas no existen de forma independiente a la medición. En este sentido, y al contrario que Penrose, adopta un enfoque cercano al de la interpretación de Copenhague, si bien con un cierto toque idealista: las propiedades cuánticas existen como estructuras que poseen la información más precisa que podemos obtener.<sup>11</sup> Para Stapp, “la información es la moneda de la realidad, no la materia: el universo es una estructura informacional, no sustantiva” (Stapp 1999, p. 149). La medición, sostiene, consiste en un intercambio de información, la cual llega a nuestro cerebro a través de los órganos sensoriales y es procesada por él. El ser humano es una conjunción de su cuerpo, su cerebro y su mente, de forma que posee un conjunto de eventos conscientes (su mente) organizados por su cuerpo y su cerebro. Tanto el observador (un ser humano) como el sistema que mide forman parte de una entidad informacional al completo, donde la reducción a un único estado cuántico del sistema medido se correlaciona con el estado cerebral del observador y con su elección acerca de qué preguntas formular al sistema. Existe una imagen física para cada evento consciente. Por lo tanto, la de Stapp “es una teoría de eventos, donde cada evento tiene un aspecto *atencional* y un aspecto *intencional*” (*ibid.*, p. 161). Además, el principio de incertidumbre de Heisenberg acerca de ciertas propiedades de las partículas (como posición, velocidad o energía) puede aplicarse a los iones calcio pre-sinápticos, haciendo impredecible la descarga de neurotransmisores. Esto hace que se cree una amplia gama de posibilidades alternativas, entre las cuales la

---

<sup>11</sup> Stapp no reconoce que su tesis sea propiamente idealista, sino que piensa que posee elementos tanto idealistas como materialistas (Stapp 1999, p. 159).

selección se realizaría a través del proceso de reducción a un único estado cuántico. Stapp niega, eso sí, que nuestros pensamientos estén exentos de constricción alguna. Más bien lo que sucede es que eventos conscientes y sucesos cuánticos forman parte de una misma estructura en la que los primeros no son reducibles a los segundos.

El neurofisiólogo John Eccles ha defendido una tesis dualista de interacción entre mente y materia, centrándose especialmente en la liberación de moléculas neurotransmisoras en las sinapsis de las neuronas del neocórtex cerebral (Eccles 1986 y 1994, Beck y Eccles 1992). Las dendritas de dichas neuronas se agrupan en grandes racimos denominados “dendrones”, y cada una de ellas termina en un botón sináptico (también llamado “terminal sináptico”). Cada botón contiene una enorme cantidad de vesículas alojando miles de neurotransmisores en su interior, los cuales pueden ser liberados a la hendidura sináptica cuando la señal eléctrica neuronal llega al terminal. La probabilidad de que esto suceda es del 25 %, y sería resultado del indeterminismo cuántico. Eccles sostiene que la mente es capaz de modificar esta probabilidad en los botones sinápticos de un dendrón al completo, afectando al funcionamiento de los mecanismos del cerebro. Con esta explicación, el indeterminismo cuántico podría generar posibilidades alternativas y el “yo” consciente, concebido como sustancia no-material, podría gobernar algunos procesos físicos del cerebro como los relativos al control de nuestras decisiones o a la creación de intenciones.

### *2.3.2. Posturas en contra*

Mientras que Jordan, Penrose, Stapp y Eccles apuestan por la correlación entre el indeterminismo cuántico y el libre albedrío, otros autores como Ted Honderich y Daniel Dennett creen que no existe dicha correlación. Honderich (1993, p. 37) sostiene que los sucesos cuánticos ocurren por azar, y no se

puede hacer responsable a nadie de decisiones cuya fuente sea de naturaleza azarosa. Para Dennett (1984, p. 77), el indeterminismo a escala cuántica no conduce al indeterminismo a niveles macroscópicos como el de las acciones humanas. Para que esto fuera así, cree que debería existir algún mecanismo con efecto amplificador que lo hiciera posible, pero también que no se ha detectado ningún mecanismo de ese tipo (véase sección 2.5.). La generación de posibilidades alternativas, que podría deberse a sucesos azarosos de origen cuántico, puede explicarse igualmente mediante la alusión a procesos aparentemente azarosos, pero que en realidad no lo son, originados por los ordenadores (*ibid.*, p. 120). Por último, aunque finalmente los sucesos cuánticos fueran causa de nuestra libertad y nuestra responsabilidad mediante un proceso de amplificación de sus indeterminaciones, no habría manera de saber si una persona puede ser hecha responsable de un determinado acto, debido a la estructura altamente compleja de nuestro cerebro (*ibid.*, p. 136).

Dennett (1991) también ha sido crítico con el interaccionismo, defendido entre otros por Karl Popper y por el ya mencionado John Eccles.<sup>12</sup> Para los interaccionistas, la materia y la mente son dos sustancias fundamentales que no son reducibles a otras y que se relacionan mutuamente. Sin embargo, esta tesis parece chocar con el principio de conservación de la energía. Mario Bunge (1980, p. 17) dice al respecto: “Si la mente inmaterial pudiera mover la materia, entonces crearía energía; y si la materia actuara sobre la mente inmaterial, entonces la energía desaparecería. En cualquier caso la energía... no sería conservada”.

---

<sup>12</sup> El interaccionismo defendido por Popper y por Eccles se enmarca dentro de la teoría de los tres mundos: el mundo 1 está formado por lo físico, el mundo 2 por lo mental y el mundo 3 por lo cultural. Esta teoría forma parte también del pensamiento de Penrose.

## 2.4. ¿Qué dice la neurobiología?

Una vía de investigación que podría resultar interesante tanto para los partidarios como para los detractores del correlato entre el indeterminismo cuántico y el libre albedrío es la relativa a la búsqueda de posibles indeterminaciones cuánticas en los procesos neurobiológicos. Weber (2005) ofrece un exhaustivo repaso de los mecanismos susceptibles de presentar este tipo de indeterminaciones. Comienza con una descripción (*ibid.*, p. 667) del mecanismo por el cual las neuronas transmiten y procesan señales (ver figura 2.2.):

Las neuronas disparan los denominados [“potenciales de acción”] a través de su fibra nerviosa o axón. Un potencial de acción es una onda de despolarización (llevada a cabo por corrientes iónicas) que se extiende a lo largo de la membrana que envuelve al axón. Un axón termina típicamente en un [determinado] número de sinapsis que conectan con otras neuronas. Cuando suficientes potenciales de acción alcanzan una sinapsis, vesículas de almacenamiento intracelular que contienen neurotransmisor son vaciadas en la hendidura que separa la membrana sináptica de la neurona vecina. El neurotransmisor se difunde rápidamente a través de esta hendidura. Cuando alcanza la membrana de la neurona vecina, se une a receptores específicos que provocan una despolarización de la membrana. El resultado es un, así llamado, [“potencial sináptico”]. Si este potencial alcanza un cierto umbral, la célula vecina dispara un nuevo potencial de acción. De esta manera, una señal puede moverse desde una neurona hasta la siguiente. Este proceso forma la base para la computación neural.

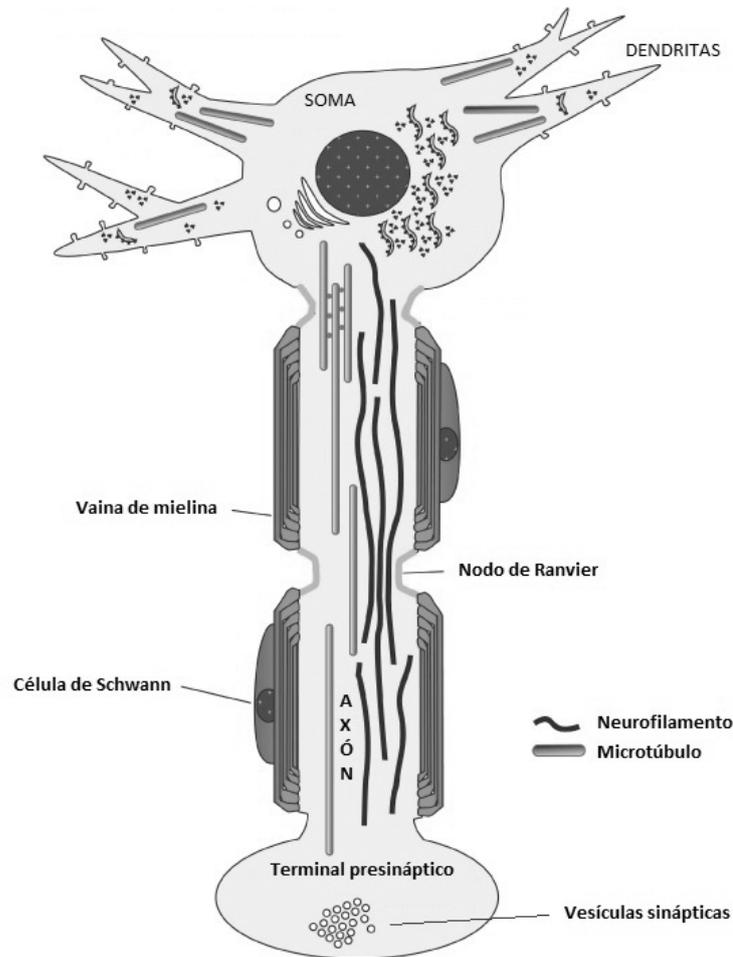


Figura 2.2. Esquema de la estructura básica de una neurona. En él se observan componentes del citoesqueleto (neurofilamentos y microtúbulos), estructuras transmisoras del potencial de acción (nodos de Ranvier, vainas de mielina y células de Schwann) y elementos de la sinapsis (terminal presináptico y vesículas sinápticas). El potencial de acción se transmite mediante saltos entre los nodos de Ranvier, dado que las vainas de mielina (producidas por las células de Schwann) actúan como aislantes eléctricos. Esto hace que el impulso nervioso se transmita más rápidamente. Modificado de Squire *et al.* (2008).

Weber continúa explicando que los sucesos aleatorios pueden estar presentes en cualquiera de las estructuras y momentos descritos en la transmisión del potencial sináptico. Además, sostiene que los procesos de medición cuántica son intrínsecamente aleatorios, y aboga por la superveniencia de lo biológico a

partir de lo físico, afirmando que la aleatoriedad a nivel macrofísico ha de estar asociada a una aleatoriedad microfísica (Weber 2005, pp. 667-8). Como consecuencia de ello, los mecanismos neurobiológicos analizados por Weber como susceptibles de presentar indeterminaciones son de naturaleza molecular. En concreto, analiza las posibilidades de tres procesos: el transporte de neurotransmisores, la difusión molecular, y la apertura y el cierre de canales iónicos.

El transporte de neurotransmisores es llevado a cabo mediante vesículas en cuyo movimiento participa el citoesqueleto celular (ver figura 2.2.). Ciertos autores (Penrose 1994, Hameroff y Penrose 1996) han sugerido que los microtúbulos (componentes del citoesqueleto) pueden manifestar coherencia cuántica cuando se encuentran en un ambiente lo suficientemente aislado, provocando un proceso indeterminista de transporte de vesículas, y presentar el fenómeno de reducción objetiva orquestada u *Orch OR* (véase apartado 2.3.1.). Sin embargo, Weber (2005, p. 668) rebate estos argumentos afirmando que no existen hallazgos experimentales que apoyen la hipótesis de coherencia cuántica en los microtúbulos,<sup>13</sup> y que estos actúan simplemente a modo de cables por los que las vesículas, con ayuda de proteínas motoras (las kinesinas y las dineínas), son conducidas hacia la hendidura sináptica, pero que no intervienen en la liberación de los neurotransmisores de las vesículas.

En cuanto al fenómeno de difusión molecular, es decir, el movimiento espontáneo de moléculas entre dos espacios con tendencia a igualar la concentración entre ambos, se sabe que tiene lugar en el momento en que los neurotransmisores atraviesan la hendidura sináptica. Weber aclara que el funcionamiento de este proceso se considera determinista actualmente (*ibid.*, pp. 668-9).

---

<sup>13</sup> Por su parte, Tegmark (2000) ha sido crítico con la relación entre coherencia cuántica y consciencia.

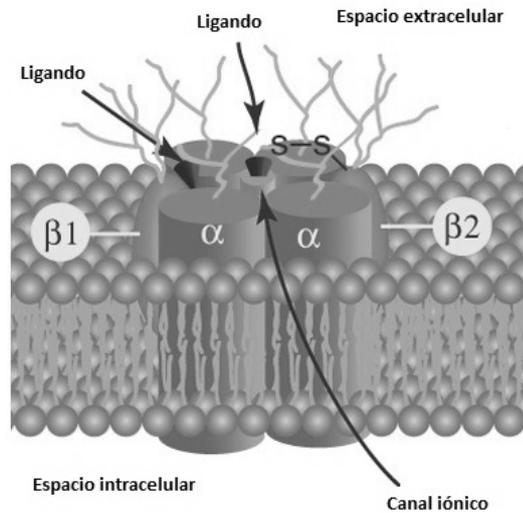


Figura 2.3. Corte transversal esquemático de una membrana celular mostrando un canal iónico de sodio. Se aprecia la división estructural del canal, formado por distintas subunidades ( $\alpha$ ,  $\beta 1$  y  $\beta 2$ ). Modificado de Squire *et al.* (2008).

El mecanismo más prometedor para la manifestación de indeterminaciones cuánticas, piensa Weber, es el de apertura y cierre de canales iónicos (ver figura 2.3.). Es por esta razón que presenta una detallada serie de argumentos a favor y en contra de esta posibilidad (Weber 2005, pp. 669-71), precedida de un párrafo explicativo acerca del funcionamiento de los canales iónicos (*ibid.*, p. 669):

El intercambio de señales entre neuronas implica esencialmente la apertura y el cierre de diferentes tipos de canales iónicos. Tales canales son moléculas proteicas comparativamente grandes que están incrustadas en la membrana neural. Son selectivamente permeables a clases específicas de iones hidratados, típicamente, iones sodio, potasio, calcio o cloruro. Los canales iónicos tienen diferentes estados, típicamente un estado de baja conductancia iónica ('cerrado'), uno de alta conductancia iónica ('abierto'), y un estado inactivo. Dependiendo del tipo específico de canal, su estado estará influenciado por la presencia de ligandos <sup>14</sup> [...] o por el voltaje a través de la

<sup>14</sup> Reciben este nombre los iones o las moléculas que forman enlaces con moléculas receptoras.

membrana. Toda la excitación eléctrica en las membranas neurales está controlada por diferentes clases de canales iónicos: [sic] Los potenciales de acción se despliegan principalmente mediante la ayuda de canales de sodio y potasio regulados por voltaje. La generación de potenciales receptores implica a canales iónicos regulados por ligandos o mecánicamente.<sup>15</sup> La liberación de neurotransmisor es iniciada por canales de Ca<sup>++</sup> [i.e., ión calcio] regulados por voltaje. Los receptores de neurotransmisor son esencialmente canales iónicos regulados por ligandos.

La posibilidad de que la apertura y el cierre de canales iónicos manifiesten indeterminaciones cuánticas, explica Weber, nace de estudios realizados mediante una técnica conocida como “patch-clamping”, y según los cuales estos canales parecen exhibir un patrón estocástico. En concreto, y fijándonos exclusivamente en una molécula, el comportamiento de un canal es irregular. Si nuestro estudio se centra, en cambio, en observar si un canal se abrirá o se cerrará, veremos que ambas opciones gozan de una probabilidad concreta, sin importar que ese mismo canal estuviera previamente abierto o cerrado. Otros estudios han mostrado, además, que lo que sucede en un canal iónico puede ascender al nivel neuronal:<sup>16</sup> se han observado potenciales de acción surgidos de modo espontáneo en cultivos celulares. Estos estudios corresponden a Johansson y Arhem (1994), así como a Chow y White (1996). Weber sugiere que “Esto suscita la posibilidad teórica de que un solo evento azaroso, quizás un evento cuántico, pueda tener alguna influencia en el comportamiento de un animal complejo” (Weber 2005, p. 670). No obstante, ofrece también argumentos y referencias en contra de esta posibilidad (Liebovitch y Toth 1991, Cavalcanti y Fontanazzi 1999), según los cuales el funcionamiento de apertura y cierre de los canales iónicos puede ser igualmente descrito por

---

<sup>15</sup> Los canales iónicos regulados mecánicamente están regulados por impulsos mecánicos, es decir, responden a acciones mecánicas. Este tipo de regulación también se presenta, por ejemplo, en las células ciliadas presentes en el oído interno, sensibles a las ondas de sonido.

<sup>16</sup> El término utilizado por Weber para describir este fenómeno es “percolate up” (2005, p. 670), algo así como “filtrar hacia arriba”.

modelos deterministas de carácter caótico cuyo desarrollo temporal es impredecible.

Los estudios y argumentos expuestos a lo largo de esta sección dejan patente la dificultad que se presenta cuando se pretende demostrar que el indeterminismo cuántico se encuentra en la base de mecanismos neurobiológicos que son claves en el comportamiento de los seres humanos, y que por tanto son susceptibles de serlo en el ejercicio de su libertad. Así lo señala también Weber: “No es suficiente con mostrar que eventos particulares a nivel molecular podrían crear la diferencia entre distintos estados macroscópicos; los indeterministas deben demostrar asimismo que estos eventos moleculares son intrínsecamente estocásticos” (Weber 2005, p. 671).

## **2.5. El caos como amplificación**

Una interesante relación física no mencionada hasta el momento es la que podría producirse entre los fenómenos de la mecánica cuántica y los descritos por la teoría del caos. Algunos trabajos han sugerido la posibilidad de un efecto amplificador del caos sobre los indeterminismos cuánticos (*e.g.*, Hobbs 1991). Los fenómenos caóticos se caracterizan por el enorme peso que las condiciones iniciales tienen en la evolución de sus modelos. Un cambio minúsculo puede tener enormes consecuencias. Se trata, no obstante, de modelos cuyo desarrollo, si bien impredecible, es de naturaleza determinista. Por todo esto, se hace necesario considerar la posibilidad de que una fluctuación cuántica indeterminista, que por sí sola podría tener efectos despreciables en lo relativo al funcionamiento macroscópico del cerebro, pueda verse amplificada y tener consecuencias sobre el comportamiento y la volición debido a la intervención del caos. Bishop (2002, p. 119) lo expresa de la siguiente forma:

[S]upongamos que las pautas de los disparos neuronales en el cerebro corresponden a estados de decisión. El caos podría amplificar los eventos cuánticos, provocando que se dispare una única neurona que de otra forma no se habría disparado. Si el cerebro (un objeto macroscópico) se encuentra también en un estado dinámico caótico, que lo hace sensible a pequeñas perturbaciones, este disparo neuronal adicional, siendo pequeño, podría ser [...] amplificado hasta el punto en que los estados cerebrales evolucionaran de forma distinta a [como lo harían] si la neurona no se hubiera disparado. A su vez estos disparos neuronales alterados y estados cerebrales podrían llevar a los efectos cuánticos al punto de afectar a los resultados de las elecciones humanas.

Tal y como el propio Bishop explica, el argumento recién expuesto no está exento de problemas, como el escaso consenso empírico acerca del funcionamiento caótico del cerebro, la existencia de interpretaciones deterministas de la mecánica cuántica <sup>17</sup> o la posibilidad de que el cerebro constriña la amplificación de los efectos cuánticos (*ibid.*, pp. 119-20).

## **2.6. Mecánica cuántica y libre albedrío: análisis crítico**

A lo largo del capítulo hemos repasado los fundamentos de la mecánica cuántica y la amenaza que esta supone para el determinismo, hemos presentado importantes interpretaciones alternativas de dicha teoría, hemos expuesto algunas posturas a favor y en contra de la correlación entre el indeterminismo cuántico y la libre voluntad, hemos ofrecido información que la neurobiología puede aportar acerca de esta hipotética correlación, y hemos destacado el posible efecto amplificador del caos sobre las indeterminaciones generadas a nivel cuántico. Llegados a este punto, es momento de emplear todos los elementos y teorías vistos en un análisis que los enmarque en el debate sobre el libre albedrío y la responsabilidad moral. Llevaremos a cabo esta tarea a partir

---

<sup>17</sup> La de Bohm es un ejemplo de este tipo de interpretaciones (véase el apartado 2.2.2.).

de las cinco cuestiones de Kane acerca de la libertad, que adelantábamos en la introducción de esta tesis: (A) Compatibilidad, (B) Significatividad, (C) Inteligibilidad, (D) Existencia, y (E) Determinismo.

### *2.6.1. Irrelevancia para la compatibilidad y la significatividad*

Los problemas de la compatibilidad y la significatividad del libre albedrío están planteados por Kane mediante las siguientes cuestiones (1996, p. 13):

[A] La Cuestión de Compatibilidad: ¿Es el libre albedrío compatible con el determinismo?

[B] La Cuestión de Significatividad: ¿Por qué queremos, o deberíamos querer, poseer un libre albedrío que es incompatible con el determinismo? ¿Es este un tipo de libertad “que valga la pena querer” (por utilizar la útil frase de Dennett)? Y si es así, ¿por qué?

Respecto a la Cuestión de Compatibilidad, pensamos que es poco lo que tiene que decir la mecánica cuántica por sí sola. No puede contribuir con datos experimentales que ayuden a esclarecer si el determinismo y el libre albedrío son compatibles, puesto que la libre voluntad, caso de existir, es susceptible de manifestarse a la escala macroscópica del cerebro (el único órgano capaz de alojar el “yo” volitivo), y la mecánica cuántica no analiza eventos cerebrales sino microfísicos. A esto hay que añadir el papel central que el indeterminismo juega en la teoría. Sea o no un indeterminismo de naturaleza ontológica, lo cierto es que una teoría indeterminista será incapaz de decir algo acerca de una hipotética compatibilidad del determinismo con la libertad.

En cuanto a la Cuestión de Significatividad, la mecánica cuántica no puede aportar nada. No nos ayuda a responder por qué los seres humanos deseamos o desearíamos tener una libertad incompatible con el determinismo. El debate sobre esta cuestión quizá pudiera ser enriquecido con aportes empíricos procedentes de la psicología y referentes a los motivos de un deseo

como el que esta cuestión plantea. Una teoría puramente física, como es la mecánica cuántica, no puede decirnos nada sobre este asunto.

Aunque, en definitiva, la mecánica cuántica resulta irrelevante para las Cuestiones de Compatibilidad y Significatividad, comprobaremos a continuación que no sucede lo mismo con las tres cuestiones restantes: de Inteligibilidad, de Existencia y Determinista. El papel de la mecánica cuántica en el debate sobre determinismo y libre albedrío depende, por tanto, del enfoque específico que adoptemos en nuestro estudio.

### *2.6.2. El control último como reto para la inteligibilidad*

Recordemos la formulación planteada para la Cuestión de Inteligibilidad (Kane 1996, p. 13):

[C] La Cuestión de Inteligibilidad: ¿Podemos darle sentido a una libertad o libre albedrío que es incompatible con el determinismo? ¿Es dicha libertad coherente o inteligible? ¿O es, como reivindican muchos críticos, esencialmente misteriosa y terminantemente oscura?

¿Es, pues, inteligible una libertad incompatibilista? Puesto que los libertaristas son los únicos incompatibilistas que creen que los humanos gozamos de libre voluntad, la pregunta podría ser formulada de esta manera: ¿Es inteligible un libre albedrío libertarista? El libertarismo, recordemos, pregona que el determinismo excluye las posibilidades alternativas y el control último y que ambas condiciones son necesarias para la libertad (véase sección 1.3.). Por lo tanto, una respuesta afirmativa a la Cuestión de Inteligibilidad que se base en la mecánica cuántica debería pasar por ofrecer una explicación que dé cuenta de cómo esta, suponiéndose indeterminista, hace posibles las posibilidades alternativas y el control último sobre nuestras decisiones y/o acciones. Es importante destacar que aquí lo importante no es la verdad de la explicación, sino solo si resulta inteligible (si podemos concebir su contenido y si es

coherente). Es por ello que no se necesita ser un libertarista para responder con un “sí” a la cuestión. Podría ser que un determinista duro respondiera en ese mismo sentido, aunque lógicamente estaría en contra de la existencia de un libre albedrío incompatible con el determinismo.

Veamos de qué manera podría la mecánica cuántica posibilitar la existencia de posibilidades alternativas. Al tratarse de una teoría basada en probabilidades (según la interpretación de Copenhague), en la que un resultado individual es impredecible, resulta sencillo imaginar que el azar cuántico pueda ocasionar alteraciones indeterministas que generen diferentes posibilidades en distintas neuronas o grupos de neuronas, constituyendo una fuente de alternativas sobre las que se pueda escoger. No sabemos si esto sucede, pero *es concebible*. Encontramos más problemas cuando abordamos la condición de control último, sobre la que seguidamente me centraré. En el apartado 2.3.1. hemos conocido cuatro importantes posturas a favor del correlato entre indeterminismo cuántico y libertad (Jordan, Penrose, Stapp y Eccles), que podrían proporcionar una fuente de argumentos a los libertaristas para apoyar su visión. Analizaremos, una a una, dichas posturas, además de la posibilidad de que el caos constituya un sistema amplificador del indeterminismo cuántico.

Según Jordan, las funciones vitales celulares estarían dominadas por los fenómenos cuánticos. Independientemente de que se trate de una opción para la cual no existe una sola evidencia empírica a favor, una libertad incompatibilista apoyada en tal afirmación resulta ininteligible. Es inconcebible el control último por parte de un agente si las funciones vitales de su cerebro se encuentran bajo el yugo del azar cuántico. Si toda función vital de la célula depende de sucesos aleatorios, no hay opción para que el “yo” volitivo organice y evalúe las informaciones transmitidas por las neuronas ni para que pueda controlar las decisiones y/o las acciones que dictamina, pues la única fuente de estas será el azar.

La reducción objetiva orquestada u *Orch OR*, descrita por Hammeroff y Penrose como un conjunto de reducciones objetivas, no-locales y

sincronizadas entre los microtúbulos de las neuronas del cerebro, no tiene el mismo problema que la propuesta de Jordan. Una libre voluntad libertarista que integre la *Orch OR* tendría posibilidades de sortear el obstáculo que supone la amenaza al control último. La coordinación entre reducciones a un único estado de un gran número de neuronas es un fenómeno no basado en propiedades de neuronas individuales, sino en sus *relaciones*. Y estas relaciones bien podrían ser un correlato de la actuación de un agente con un poder causal de tipo no-local. No se trataría de que cada neurona individual produjera un efecto cuántico decisivo en las decisiones y/o acciones del agente. Por el contrario, la *Orch OR* sería una manifestación del pensamiento consciente del agente, no su causa, lo que abriría la posibilidad de que parte de ese pensamiento consciente fuera una manifestación de su libre voluntad. Una libertad que integrara esta descripción tendría naturaleza indeterminista pero no vería amenazado su componente de control último; sería, por tanto, inteligible.

Stapp cree que la consciencia se sustenta en la existencia de una entidad informacional formada por un sistema cuántico y su observador, en la que la reducción a un único estado cuántico del primero se correlaciona con el estado cerebral del segundo. Considera que la clave es el intercambio de información, no la materia. Sin embargo, un libre albedrío libertarista que se sustente en esta explicación es difícilmente concebible. Tal y como Stapp afirma, los pensamientos no están libres de constricción alguna, pero a su vez no son reducibles a sucesos microfísicos sino que forman parte de una estructura informacional más amplia. Esto hace imposible situar el lugar donde podría situarse el control causal dentro del dúo observador – sistema cuántico. No parece haber una direccionalidad causal, sino más bien un sistema de flujo de información en el que ninguna de las partes es causa de la otra. Dadas estas circunstancias, la hipótesis de Stapp no es prometedora como elemento de una explicación para una libertad incompatibilista, pues no puede dar cuenta adecuadamente de su condición de control último.

La hipótesis de Eccles, centrada en la capacidad de la mente para modificar la probabilidad de liberación de neurotransmisores en los botones presinápticos de los dendrones, afectando al funcionamiento de los mecanismos del cerebro, podría ofrecernos la posibilidad de que el “yo” consciente controle ciertos procesos cerebrales como la formación de intenciones o la toma de decisiones. Sin embargo, un elemento nuclear de esta hipótesis es el convencimiento de que este “yo” consciente es una sustancia inmaterial que interacciona con la materia. Una tesis dualista de este tipo no ayuda a completar y hacer inteligible una explicación libertarista para el libre albedrío, sino que, por el contrario, añade nuevos conflictos, como, por ejemplo, la naturaleza de la conexión entre esa sustancia inmaterial y el sistema nervioso. Además, y en relación con el segundo principio de la termodinámica, no sabemos cómo podría evitar esta hipótesis el problema que supone la creación de nueva energía en un sistema en el que una mente inmaterial tiene un efecto causal sobre la materia, como bien apuntaba Bunge.

Dejando a un lado las propuestas concretas de los autores recién mencionados, vale la pena comentar la posibilidad de que, tal y como recuerda Bishop, el fenómeno del caos, de naturaleza impredecible pero determinista, tenga un efecto amplificador sobre la mecánica cuántica de forma tal que sus indeterminaciones pudieran ascender a niveles macroscópicos. Es concebible que indeterminaciones puntuales puedan trepar desde los niveles microfísicos de escala cuántica hasta los niveles macrofísicos del cerebro, incrementándose sus efectos. Esto podría eventualmente generar posibilidades alternativas. Sin embargo, no es inteligible la forma en que esta interacción cuántico-caótica podría posibilitar la condición de control último. ¿Cómo podría un agente poseer el control sobre sus decisiones si el origen de estas depende, en último término y exclusivamente, de eventos aleatorios a escala microfísica? Una teoría libertarista del libre albedrío, si pretende ser inteligible, no puede confiar en el efecto amplificador del caos sobre los eventos cuánticos como única explicación.

Como podemos ver, resulta altamente problemático explicar la condición de control último en un libre albedrío libertarista basándonos en la mecánica cuántica. La única hipótesis susceptible de ser integrada en una explicación libertarista, de forma que esta resulte inteligible, es la ofrecida por Hammeroff y Penrose. Las propuestas de Jordan, Stapp y Eccles tienen enormes dificultades si se pretenden utilizar como elementos de una teoría perteneciente al libertarismo. Por otro lado, cuando Honderich sostiene que no podemos hacer responsable a nadie de decisiones cuya fuente sea de naturaleza azarosa, como podría ser el caso de los sucesos cuánticos, es posible que piense en la dificultad para explicar el control último. No obstante, la mecánica cuántica podría ayudar a comprender el surgimiento de posibilidades alternativas y ligarse a explicaciones para el control último no basadas en ella. En este sentido, no hay motivo para rechazar que un agente pueda ser responsable de decisiones en cuyo proceso de formación interviene la mecánica cuántica.

### *2.6.3. Indicios contra la existencia*

La hipotética existencia de un libre albedrío libertarista queda planteada de esta forma (Kane 1996, p. 13):

[D] La Cuestión de Existencia: ¿Dicha libertad existe en realidad en el orden natural? Y si es así, ¿dónde?

La Cuestión de Inteligibilidad, planteada en el apartado anterior, apelaba a la coherencia de las explicaciones libertaristas, sin entrar a valorar la existencia o inexistencia de las libertades que proclaman. Esta tarea es solicitada en la Cuestión de Existencia, la cual podría quedar formulada de la siguiente manera si la contextualizamos en el tema del presente capítulo: “¿Existe en el orden natural una libertad incompatible con el determinismo y compatible con la

mecánica cuántica?”. Hasta la fecha no se ha producido un hallazgo en sentido afirmativo, pero sí podemos analizar los indicios empíricos que existen al respecto.

La hipótesis de Jordan es, dado nuestro conocimiento actual, altamente improbable. No hay prueba alguna a favor de un control absoluto de las funciones vitales de nuestras células por parte de los eventos cuánticos. Se conoce una amplia gama de mecanismos de control neuronal altamente organizados y con base macroscópica para los cuales es inexplicable un origen cuántico. A día de hoy podemos calificar de imposible la existencia de una libertad sustentada en la propuesta de Jordan.

Las hipótesis de Stapp y Eccles, si bien son susceptibles de ser contrastadas experimentalmente en alguno de sus elementos, como la probabilidad de descarga de neurotransmisores, son imposibles de comprobar en otros componentes que son claves para su articulación. En el caso de la estructura informacional de Stapp hay una reducción a un único estado cuántico del sistema medido correlacionada con el estado cerebral del observador y su elección acerca de qué preguntas formular al sistema. El problema de esta explicación es que no tenemos forma de distinguir si el estado cerebral del observador nace de la reducción a un único estado o si dicha reducción surge del estado cerebral del observador. En cuanto a la propuesta de Eccles, el problema radica en su carácter interaccionista. Si bien este autor ofrece una detallada descripción anatómica y funcional de su hipótesis, no hay evidencia empírica de la existencia de una mente inmaterial que tenga efecto sobre el cerebro material. Y es lógico que así sea, pues una sustancia inmaterial no puede ser detectada con técnicas e instrumentos materiales como los utilizados por los investigadores experimentales. Abogar por una mente no material es una postura que no puede ser corroborada por la ciencia actual. En definitiva, no hay pruebas empíricas que sustenten la posible existencia de un tipo de libertad incompatibilista basado en la hipótesis de Stapp o en la de Eccles.

La hipótesis de reducción objetiva orquestada, descrita por Hammeroff y Penrose, tiene la ventaja de estar en gran parte compuesta por elementos susceptibles de ser estudiados experimentalmente. Al situar el proceso *Orch OR* en los microtúbulos de las neuronas cerebrales, estos autores nos ofrecen, al contrario que los anteriormente mencionados, un lugar concreto donde buscar. Para apoyar la hipótesis *Orch OR* necesitaríamos encontrar un conjunto de reducciones objetivas, no-locales y sincronizadas en los microtúbulos, que provocaría un transporte indeterminista de las vesículas sinápticas. Sin embargo, y como ya dijimos, Weber señala que no existen actualmente hallazgos experimentales que apoyen la coherencia cuántica que necesitan los microtúbulos para que se produzca la reducción objetiva orquestada, pero que, en cambio, sí parece haber indicios que apuntan a que los microtúbulos actuarían meramente como cables por los que las vesículas que contienen moléculas de neurotransmisor se dirigen hacia la hendidura sináptica. Es por ello que el modelo *Orch OR*, a día de hoy, no goza de pruebas concluyentes a su favor, de manera que no podemos afirmar que exista una libre voluntad libertarista que lo integre.

La posible amplificación de las indeterminaciones cuánticas debida a la actuación del caos también está sujeta a la Cuestión de Existencia. A este respecto, Bishop destaca que existe escaso consenso empírico sobre el posible funcionamiento caótico del cerebro, y que no debemos olvidar la posibilidad de que este constriña la amplificación. Dennett apunta hacia la necesidad de un mecanismo amplificador para que el indeterminismo a escala cuántica pueda conducir al indeterminismo al nivel macroscópico de las acciones humanas, pero destaca que no se ha detectado mecanismo alguno de esta clase. Si atendemos a afirmaciones como las de estos dos autores, no es posible sostener la existencia de una clase de libertad incompatibilista basada en la amplificación caótica de indeterminaciones cuánticas en nuestros cerebros.

A la vista de lo expuesto en este apartado, no podemos ofrecer una respuesta afirmativa a la Cuestión de Existencia, puesto que las principales

hipótesis que podrían ser utilizadas para ello están obsoletas (Jordan), son poco susceptibles a la comprobación empírica (Stapp, Eccles) o no gozan de apoyo experimental suficiente (Hammeroff y Penrose, efecto amplificador del caos). No obstante, cabría profundizar en el estudio de estas últimas por su contenido altamente comprobable desde el punto de vista empírico.

#### *2.6.4. Dudas sobre el determinismo*

El problema de la verdad del determinismo como ley universal queda planteado del siguiente modo (Kane 2002a, p. 6):

[E] La Cuestión Determinista: ¿Es el determinismo cierto?

Si obtuviéramos un “no” a esta cuestión basándonos en la mecánica cuántica, la solución sería definitiva: un solo caso de indeterminismo es suficiente para rechazar el determinismo como ley universal. Pero no sucede igual a la inversa: aunque concluyéramos que todos los fenómenos cuánticos son deterministas, no deberíamos olvidar la posibilidad de encontrar indeterminaciones en otros ámbitos. El objetivo en este apartado es discutir de qué modo puede ayudar la mecánica cuántica a responder al problema del determinismo. A lo largo de las secciones 2.1. y 2.2. presentamos las características principales tanto de la interpretación más extendida de la mecánica cuántica (la de Copenhague) como de algunas interpretaciones alternativas. En el apartado 2.1.7. destacamos la posibilidad de interpretar el indeterminismo cuántico de dos modos: ontológico o epistemológico. A la vista de todas estas opciones, ¿qué podemos decir sobre si el determinismo es cierto o no a nivel cuántico?

Si entendemos el determinismo como la doctrina según la cual sólo puede existir un único futuro, dados un pasado y unas leyes de la naturaleza concretas, entonces el principio de incertidumbre (incapacidad de averiguar simultáneamente el valor de dos observables de una partícula) y el problema de

la medida (superposición de estados que sólo se resuelve cuando se realiza una medición) pueden suponer una seria amenaza bajo la interpretación de Copenhague. Los partidarios de dicha interpretación consideran que el principio de incertidumbre no refleja nuestra incapacidad para ahondar en los secretos de la materia, sino que es una propiedad intrínseca de la naturaleza. Creen, además, que la lectura correcta del problema de la medida consiste en considerar que la superposición de estados previa a la medición es real, lo que conduce a conclusiones, a nivel microfísico, similares a la conclusión de la paradoja del “gato de Schrödinger” (en ausencia de medición, el gato estaría vivo y muerto al mismo tiempo). Por otro lado, en la interpretación de Bohm las partículas poseen propiedades definidas independientemente de si son medidas o no, y la impredecibilidad acerca de estas propiedades nace de los efectos de un campo cuántico. Esta interpretación, al contrario que la anterior, es determinista desde el punto de vista ontológico, si bien acepta el indeterminismo epistemológico de los fenómenos cuánticos.

En el apartado 2.2.1. expusimos dos interpretaciones más. La suma de historias posibles de Feynman (la probabilidad de que una partícula haga un recorrido entre dos puntos se averigua sumando todas las trayectorias posibles entre ambos) y la interpretación de los mundos múltiples (múltiples observaciones que tienen lugar simultáneamente y que no saben nada unas acerca de las otras) pueden parecer ontológicamente deterministas a simple vista (muchos autores así lo consideran), pues eliminan la impredecibilidad de los sucesos cuánticos al considerar que ocurre un enorme número de ellos al mismo tiempo. Sin embargo, esta simultaneidad contradice la existencia de un único futuro, necesaria para aceptar el determinismo definido por van Inwagen (1983, p. 65) de la siguiente manera: “El determinismo, es, intuitivamente, la tesis según la cual, dados el pasado y las leyes de la naturaleza, existe un solo futuro posible”. El determinismo (al menos entendido de esta manera) no es posible si varios futuros tienen lugar al mismo tiempo.

A nuestro modo de ver, no existen motivos suficientes para decantarnos por alguna de estas interpretaciones en cuanto a su determinismo o indeterminismo se refiere. En primer lugar, la interpretación de la suma de historias posibles se ha demostrado eficaz como instrumento a nivel teórico, pero no se han encontrado indicios experimentales a su favor. En segundo lugar, la interpretación de los mundos múltiples no puede ser comprobada experimentalmente (al menos en la actualidad), pues prevé que las distintas observaciones que tienen lugar al mismo tiempo no se conocen entre sí. No tenemos forma de saber si la observación real que experimentamos tiene lugar, o no, al mismo tiempo que otras. En tercer lugar, la interpretación de Bohm es ciertamente interesante como alternativa a la interpretación de Copenhague, y lo es por su coherencia y por su capacidad para explicar fenómenos también integrados por aquella, como la no-localidad y las predicciones basadas en probabilidades. No obstante, no hay evidencia experimental a favor de la existencia del campo cuántico descrito por Bohm. En cuarto y último lugar, tampoco existe un motivo irrenunciable que nos impulse a pensar que el principio de incertidumbre integrado en la interpretación de Copenhague refleja un indeterminismo ontológico de la mecánica cuántica, por las razones que expondremos a continuación.

Como dijimos en el apartado 2.1.4., cuando queremos averiguar al mismo tiempo la posición y el momento lineal de una partícula nos encontramos con una seria dificultad. Uno de los modos que tenemos de intentarlo consiste en iluminar dicha partícula para saber dónde se encuentra. El problema surge porque si iluminamos con luz de corta longitud de onda, la partícula verá alterado su momento lineal al rebotar la luz en ella (aunque conoceremos su posición con precisión), y si iluminamos con luz de alta longitud de onda, tendremos un gran margen de error a la hora de averiguar su posición (aunque el momento lineal no se verá muy alterado y lo podremos conocer con precisión). Esta forma de estudiar una partícula refleja la ineficacia del método a la hora de analizar objetivamente el sistema en estudio, porque el

método influye en el resultado. ¿Por qué debemos suponer que, previamente a una medición que se muestra ineficaz, el sistema cuántico no posee propiedades intrínsecas? De un modo similar, podríamos encontrar ante esta situación con otros tipos de métodos y de observables. La afirmación de que no tenemos razones para creer que las propiedades estaban presentes antes de realizar la medición es obviamente correcta: no podemos saber si algo está ahí hasta que lo percibimos. Pero tampoco deja de ser cierta la afirmación contraria: *no hay razones para rechazar* que las propiedades ya estaban ahí antes de que nosotros las midiéramos.<sup>18</sup> Frente a esto podría esgrimirse el célebre Teorema de Bell (1964), según el cual no es posible la reproducción de las predicciones mecánico-cuánticas por medio de una teoría de variables ocultas de carácter local. Sin embargo, debe aclararse que este teorema nos obliga a rechazar: 1) el realismo (*i.e.*, que las propiedades del sistema cuántico existen previamente a su medición), 2) el principio de localidad, o 3) ambos a la vez. Esto deja abierta la posibilidad de una teoría realista no-local, como la interpretación de Bohm. Ahora bien, las teorías de variables ocultas, y entre ellas la de Bohm, quedan seriamente dañadas a la luz del teorema Kochen – Specker (Kochen y Specker 1967). Este teorema demuestra que en un sistema cuántico no es posible atribuir valores a *todas* sus propiedades observables con precisión y simultaneidad. Una atribución precisa y simultánea resulta alcanzable cuando se limita a las propiedades observables relativas a un contexto de medición específico. A este respecto, Lombardi, Castagnino y Ardenghi (2009, p. 9) dicen que “todo intento de adjudicar valores actuales precisos a todos los observables del sistema, de modo tal que las probabilidades puedan interpretarse por ignorancia respecto de un estado subyacente a la manera de la mecánica estadística clásica, conduce a contradicción”. No obstante, y a pesar de su potencia, el teorema Kochen – Specker no ha podido ser comprobado empíricamente. Además, hay posibles

---

<sup>18</sup> Las teorías de colapso objetivo, sin ir más lejos, sostienen que se da una destrucción de la superposición de estados cuánticos incluso en ausencia de medición

escapatorias que, por el momento, impiden que se haya llevado a cabo un experimento que, de un modo definitivo, confirme el teorema de Bell (Cabello 2007, p. 60). Por todo ello, pensamos que no puede afirmarse que alguna de las interpretaciones de la mecánica cuántica demuestre definitivamente, desde un punto de vista ontológico, que el indeterminismo cuántico sea cierto o que no lo sea. Consideramos, en sintonía con otros autores (*e.g.*, Balaguer 2010, p. 147), que el debate al respecto es inconcluyente. En este sentido, cabe destacar que durante los últimos años han aparecido trabajos relevantes con opiniones muy distintas respecto al supuesto indeterminismo cuántico. Tanto Gerard 't Hooft (2007) como otros autores (*e.g.*: Sadiq y Ghafoor 2010, Hofmann 2012, Brassard y Raymond-Robychaud 2013) han mantenido posturas deterministas. Por el contrario, Conway y Kochen (2006, 2009) han construido “el Teorema del Libre Albedrío”, mediante el cual pretenden demostrar la indeterminación en los fenómenos cuánticos. Este teorema ha suscitado un encendido debate (véanse Tumulka 2007, Goldstein *et al.* 2010).

Antes de finalizar con la discusión acerca de la Cuestión Determinista, vale la pena recordar unas observaciones acerca del indeterminismo cuántico en relación con la neurobiología. En la sección 2.4. expusimos, siguiendo las directrices utilizadas por Weber (2005), tres procesos neurobiológicos moleculares susceptibles de presentar indeterminaciones cuánticas: el transporte de neurotransmisores, la difusión molecular, y la apertura y el cierre de canales iónicos. En primer lugar, el transporte de neurotransmisores necesita de una coherencia cuántica para la cual no pueden presentarse hallazgos experimentales a su favor. En segundo lugar, la difusión molecular es tratada actualmente como un proceso determinista. En tercer lugar, para la apertura y el cierre de canales iónicos existen estudios que muestran un patrón estocástico del proceso, pero también se han confeccionado modelos deterministas que pueden describirlo. Por lo tanto, podemos decir que el indeterminismo cuántico es incierto al nivel de los fenómenos moleculares de

la neurobiología, si bien debería profundizarse en su posible manifestación en los procesos de apertura y cierre de canales iónicos.

En conclusión: ciñéndonos a la mecánica cuántica, ¿es el determinismo cierto? Teniendo en cuenta los datos actuales de los que disponemos, y a la vista de las argumentaciones contenidas en este apartado, todo lo que podemos responder a esta Cuestión Determinista es: “no lo sabemos”. Se trata de una cuestión fundamental sobre la que la ciencia tiene mucho que decir y, por tanto, debe seguir indagando.

## Capítulo 3. Explorando el nivel “macro”: neurofilosofía y libre albedrío

«Yo no me marchó sin tí, Mac...  
No podría dejarte en este estado...  
Tú vendrás conmigo...»

*Alguien voló sobre el nido del cuco* (Milos Forman, 1975).<sup>1</sup>

En la década de los 80, Patricia Churchland publicó su influyente obra *Neurophilosophy: Toward a Unified Science of the Mind/Brain* (Churchland 1986). Desde entonces, ha sido fecundo el trabajo que números filósofos y científicos han realizado en torno a las repercusiones que los descubrimientos de la neurociencia tienen para la filosofía, especialmente para la filosofía de la mente. En consecuencia, la neurofilosofía constituye un campo fundamental para el estudio del libre albedrío.

---

<sup>1</sup> El guión de esta película es obra de Bo Goldman y Lawrence Hauben, y se basó en la novela homónima de Ken Kesey. El fragmento transcrito corresponde a la versión doblada al castellano.

### 3.1. La neurofilosofía y la búsqueda de un nexo entre el cerebro y la mente

La neurofilosofía es la disciplina que analiza la repercusión filosófica de los descubrimientos de la neurociencia. Puede que el tema más importante que la neurofilosofía pretende abordar sea el problema mente-cuerpo, que nos invita a preguntarnos qué relación, si es que la hay, guardan los procesos nerviosos y los mentales. Estamos ante un problema clásico de la filosofía, y desde luego afecta directamente al debate sobre el libre albedrío si pretendemos estudiarlo desde una perspectiva neurofilosófica. Por esta razón, y para contextualizar adecuadamente nuestra aproximación a la relación de los fenómenos cerebrales con el libre albedrío, sería preceptivo comenzar ofreciendo una panorámica de los principales enfoques existentes acerca del problema mente-cuerpo. Presentaremos a continuación, y con mero propósito introductorio, un pequeño listado modificado a partir de una clasificación más profundizada ofrecida por Walter (2001, pp. 87-123).

Posiciones dualistas → Cerebro y mente son dos sustancias distintas.

- *Paralelismo* → Si bien el cerebro y la mente son mutuamente independientes, ambos se sincronizan de alguna manera. Podríamos destacar tres variantes:
  - *Autonomismo* → La sincronización entre cerebro y mente sucede de modo aleatorio.
  - *Ocasionalismo* → Dios es responsable de controlar la sincronización de forma continua.
  - *Armonía preestablecida* → Dios estableció para siempre la sincronización cuando sucedió la creación.
- *Epifenomenismo dualista* → Mientras que el cerebro controla los fenómenos mentales, lo mental carece de control sobre el cerebro.

- *Animismo* → Hay un alma que habita en toda materia y la controla. Concretando en la relación entre mente y cerebro, la primera controlaría al segundo, pero no viceversa.
- *Interaccionismo* → Cerebro y mente interactúan mutuamente de modo que cualquiera de los dos puede actuar sobre el otro.

Posiciones monistas → Sólo existe una sustancia.

- *Espiritualismo* → Lo único que existe es la mente o el espíritu, y la materia es una mera percepción mental o espiritual.
- *Materialismo estricto* → La materia es lo único que existe.
- *Monismo neutral* → El cerebro y la mente no son otra cosa que manifestaciones distintas de una misma sustancia.
- *Conductismo filosófico* → Los estados y procesos mentales pueden describirse en términos de la conducta y sus disposiciones.
- *Epifenomenismo monista* → La mente es un epifenómeno del funcionamiento del cerebro, y no ejerce influencia causal alguna sobre este.
- *Materialismo eliminativo* → Los estados y procesos mentales no existen, y deberíamos descartar las expresiones mentales en favor de las de la neurociencia.
- *Teoría de la identidad* → Los estados y procesos mentales son idénticos a los estados y procesos nerviosos. Destacan tres variantes:
  - *Teoría de la identidad de tipo* → Cada tipo (clase) de estado o proceso mental es idéntico a un tipo de estado o proceso nervioso.
  - *Teoría de la identidad de casos* → Cada caso concreto de estado o proceso mental es idéntico a un caso concreto de estado o proceso nervioso, y no con un tipo.
  - *Teoría de la identidad de rol causal* → Los estados y procesos mentales juegan un papel (rol) de tipo causal y son idénticos a estados y

procesos nerviosos que desempeñen el mismo rol causal. Esta identidad, además, puede ser diferente entre especies o incluso entre personas distintas.

- *Funcionalismo* → Independientemente de su correlato nervioso, los estados y procesos mentales se agrupan según su función: son estados y procesos funcionales.
- *Materialismo no reductivo* → Los estados y procesos mentales tienen un correlato nervioso, pero también es cierto que no es posible reducirlos a estados y procesos nerviosos. Podemos destacar tres variantes:
  - *Monismo anómalo* → Se basa en tres principios. En primer lugar, existe una relación de tipo causal entre casos mentales y estados o procesos nerviosos. En segundo lugar, esta relación causal está sujeta a una estricta ley natural. Y en tercer lugar, no existen leyes de índole psicológica ni psicofísica. La conclusión es que no pueden ser predichos los estados y procesos mentales.
  - *Dualismo de propiedades* → Aunque pertenecen a la misma sustancia, los estados y procesos físicos poseen propiedades materiales y los estados y procesos mentales poseen propiedades no materiales, y ambos tipos de propiedades no son reducibles entre sí.
  - *Emergentismo* → Las propiedades mentales emergen a partir de las propiedades nerviosas y, sin embargo, no son reducibles a ellas.
- *Instrumentalismo* → Los estados y procesos mentales no son más que un instrumento abstracto que utilizamos cuando estudiamos el conocimiento y el comportamiento, y su uso se justifica por su utilidad para la predicción o para el avance en nuestras investigaciones.
- *Pragmatismo* → No podemos afirmar la realidad de aquello que escapa a nuestros límites experimentales y a nuestros esquemas conceptuales humanos, y por tanto debemos estudiar los estados y procesos mentales en función de criterios que atiendan a la practicidad.

La lista precedente no pretende ser exhaustiva, ni tampoco introducirnos en una discusión sobre qué posturas acerca del problema mente-cuerpo son más adecuadas, si bien es verdad que algunas posiciones están actualmente más en boga que otras, e incluso las hay que están consideradas como obsoletas. Además, algunas de ellas no pueden ser contrastadas empíricamente con la ayuda de la neurociencia. Es el caso del paralelismo, al animismo y al espiritualismo. Obviaremos dichas posiciones de aquí en adelante, puesto que el objeto de estudio de este capítulo es la aportación de la neurociencia a la filosofía del libre albedrío. Por otra parte, hay que tener en cuenta que existen relaciones complejas entre algunas de las posturas de la lista. Sin ir más lejos, el monismo anómalo de Davidson podría considerarse como una teoría de la identidad de casos. En cualquier caso, el objetivo no es profundizar en los detalles de las diferentes teorías expuestas. No obstante, y como bien observa Walter (*ibid.*, p. 127), parece que si queremos utilizar la neurofilosofía para aproximarnos al problema del libre albedrío, necesitaríamos antes tener una respuesta clara acerca del problema mente-cuerpo. Sin embargo, este problema viene siendo protagonista de encendidos debates desde hace siglos y parece estar lejos de su resolución. Ante esta situación cabe preguntarse: ¿podríamos abordar la neurofilosofía del libre albedrío sin necesidad de situarnos en alguna de las posiciones que se han mantenido a lo largo del inconcluso debate sobre el problema mente-cuerpo? Podríamos dar una respuesta positiva a esta pregunta si nos inspiramos en un método ofrecido por el propio Walter (*ibid.*, pp. 127- 34): la neurofilosofía mínima.

### **3.2. El enfoque de la neurofilosofía mínima**

Tal y como observa Walter, da la sensación de que son muchos los rasgos que diferencian a las distintas posiciones neurofilosóficas (Walter 2001, p. 127), pero también es cierto que “de alguna manera todas las neurofilosofías han

contribuido con material importante para resolver ciertos puzzles” (*ibid.*, p. 128). Su propuesta, llamada “neurofilosofía mínima”, trata de poner de relieve las características comunes entre ellas e integrarlas lo máximo posible. Para apoyar esta manera de enfocar la neurofilosofía, Walter ofrece tres razones (*ibid.*, pp. 128-31). La primera es que existen temas de índole metafísica de los cuales no se espera una solución de manera definitiva, pero que sin embargo se relacionan con afirmaciones que con frecuencia corresponden a premisas infundadas. La segunda razón es la existencia de problemas neurofilosóficos concretos de suma importancia que no pueden demorarse hasta una hipotética resolución de las cuestiones sobre el problema de la relación entre mente y cerebro. Y la tercera razón es que no podemos presuponer, como hacen las teorías mente-cuerpo, que todos los tipos distintos de estados mentales se ajustan a la misma explicación metafísica. Walter subraya (*ibid.*, p. 130): “¿Por qué los estados o procesos mentales como pensar, sentir y decidir deberían estar todos relacionados con los estados físicos de la misma manera? Por supuesto es más elegante explicar *todos* los fenómenos mentales con una *sola* teoría. Pero he sido incapaz de encontrar un solo argumento convincente sobre por qué debería ser esto así”.

En relación con la tercera razón, Walter explica que podemos encontrar estados mentales muy distintos (*ibid.*, pp. 129-30): dolor, imaginación, deseo, creencia, enfado, placer, decisión, intención... y un largo etcétera. A día de hoy no podemos saber qué explicación metafísica se ajusta a cada estado mental, de forma que el autor propone seguir una “tesis de metafísica diferencial” (*ibid.*, p. 130): suponer que cada estado mental puede ajustarse a una explicación metafísica particular y no necesariamente igual a la que se ajustan otros estados.<sup>2</sup> En este sentido, expone una lista de las posibles explicaciones (*ibid.*, p. 131):

---

<sup>2</sup> Walter recomienda, no obstante, adoptar a efectos prácticos una sola relación metafísica como “opción *por defecto*” para afrontar problemas concretos (Walter 2001, p. 134). La relación más adecuada, en su opinión, es la de superveniencia (*ibid.*, pp. 134-5).

Algunos procesos <sup>3</sup> mentales podrían ser

- en principio no explicables,
- procesos de una sustancia mental,
- correlacionados uno a uno con procesos cerebrales,
- procesos emergentes,
- supervenientes sobre procesos físicos,
- mereológicamente supervenientes sobre procesos físicos,
- entendidos como roles causales de procesos físicos,
- idénticos en tipo a procesos cerebrales,
- idénticos en caso a procesos cerebrales,
- descripciones abstractas de procesos cerebrales,
- constructos interpretacionales lingüísticos,
- constructos cerebrales, o
- inexistentes

Walter aclara que la neurofilosofía mínima es “neuro” porque se centra en el estudio de los procesos mentales que presentan los seres vivos dotados de cerebro, y que es “mínima” porque aspira a utilizar un número de supuestos metafísicos que sea lo más pequeño posible (*id.*). Siguiendo estas dos orientaciones, expone las tres tesis que guardan en común las distintas neurofilosofías, que pueden ser consideradas, por tanto, como las tres tesis nucleares de la neurofilosofía mínima (*ibid.*, pp. 131-2):

(T1) *Ontología*: Los procesos mentales de los organismos biológicos son llevados a cabo por o con la ayuda de procesos neuronales.

(T2) *Constricción*: El análisis filosófico de procesos mentales no debería contradecir las mejores teorías del cerebro disponibles actualmente.

(T3) *Principio Heurístico*: El conocimiento acerca de la estructura y la dinámica de los procesos mentales puede ser adquirido a partir del conocimiento acerca de la estructura y la dinámica de los procesos neuronales.

---

<sup>3</sup> A lo largo de las pp. 130-31, Walter utiliza de forma confusa las expresiones “estado mental” y “proceso mental”, dándoles un significado que parece idéntico. No obstante, la palabra “estado” se refiere a una situación del aquí y del ahora, mientras que el significado de la palabra “proceso” está vinculado con el movimiento y con el paso del tiempo. El propio Walter realiza una distinción un poco más adelante y se decanta por la palabra “proceso” en detrimento de “estado”, “propiedad” y “evento”(pp. 131-2).

Walter destaca que (T1) parecería ajustarse únicamente a posiciones monistas (*ibid.*, p. 132). Sin embargo, sostiene que el dualismo podría encajar en ella de dos maneras: o bien sosteniendo que el cerebro es el lugar donde son llevados a término los estados mentales, o bien afirmando que el cerebro sirve de ayuda a la mente autoconsciente (como sustancia distinta) para producir estados mentales (*id.*). También el materialismo eliminativo, dice el autor, parece entrar en conflicto con (T1) (*ibid.*, p. 133). No obstante, añade, lo que los partidarios de esta postura sostienen es la inexistencia de estados mentales entendidos como constructos de la psicología popular, con lo cual podrían considerar que los procesos mentales mencionados en (T1) no hacen referencia a ellos, sino a procesos con existencia real (*id.*). Walter también subraya que si atendemos a (T2) podríamos pensar que no se ajusta a los postulados del dualismo interaccionista, pero destaca que los partidarios de esta corriente tratan actualmente de encontrar una interacción de la mente autoconsciente con el mundo físico (*id.*).

Centrémonos ahora en (T3). Debemos observar que se trata de una tesis que exige, al menos en parte, una traducción entre lo neuronal y lo mental. Hay, por tanto, un posicionamiento metafísico del autor. El propio Walter admite que “[e]l significado de (T3) se vuelve aparente cuando examinamos las metafísicas funcionalista y paralelista” (*id.*). No obstante, piensa que si bien es verdad que las propiedades neurofisiológicas podrían no constituir la única alternativa para que puedan producirse los estados mentales, es igual de cierto que sólo se sabe que posean estados mentales el hombre y algunos otros animales dotados de cerebro (*ibid.*, pp. 133-4). Actualmente, “no sabemos si otros sistemas además de los humanos pueden también exhibir propiedades mentales, y la neurofilosofía mínima solamente pretende desarrollar una teoría de los estados mentales en organismos biológicos” (*ibid.*, p. 134).

La neurofilosofía mínima, en definitiva, parece ser un instrumento de trabajo útil a la hora de estudiar las repercusiones que tienen los

descubrimientos de la neurociencia para el debate sobre el libre albedrío,<sup>4</sup> pues sus tres tesis centrales se complementan muy bien con la metodología científica de la neurociencia y, además, no exigen elegir una posición concreta en el problema mente-cuerpo entre las que otorgan un papel importante a las contribuciones de la neurociencia. En consecuencia, a lo largo de lo que resta de capítulo nos inspiraremos en la aproximación que nos brinda la neurofilosofía mínima para estudiar de qué manera podrían ser importantes dichas contribuciones para el debate sobre el libre albedrío.

### **3.3. Bases biológicas de la volición**

#### *3.3.1. Un poco de neuroanatomía*

Para poder comprender mejor el contenido de la sección, se hace conveniente comenzar con una breve descripción neuroanatómica.

---

<sup>4</sup> Walter utiliza el enfoque de la neurofilosofía mínima como herramienta para desarrollar su propia teoría neurofilosófica del libre albedrío. En su opinión, el concepto libertarista de libre albedrío no es apropiado y debe ser sustituido por el concepto “autonomía natural” (Walter 2001, p. 292). De modo abreviado, Walter explica la autonomía natural de la siguiente manera (*ibid.*, p. 299):

Poseemos autonomía natural cuando bajo circunstancias muy similares podríamos también hacer algo distinto de lo que realmente hacemos (debido a la naturaleza caótica de nuestro cerebro). Esta elección es comprensible (inteligible – está determinada por eventos pasados, por procesos de adaptación inmediata en el cerebro, y parcialmente por nuestro ambiente lingüísticamente formado), y es auténtica (cuando a través de circuitos de reflexión con ajustes emocionales podemos identificarnos con esa acción). Esta clase de autonomía es apropiada para un concepto compatibilista de responsabilidad y lo complementa en algunas áreas.

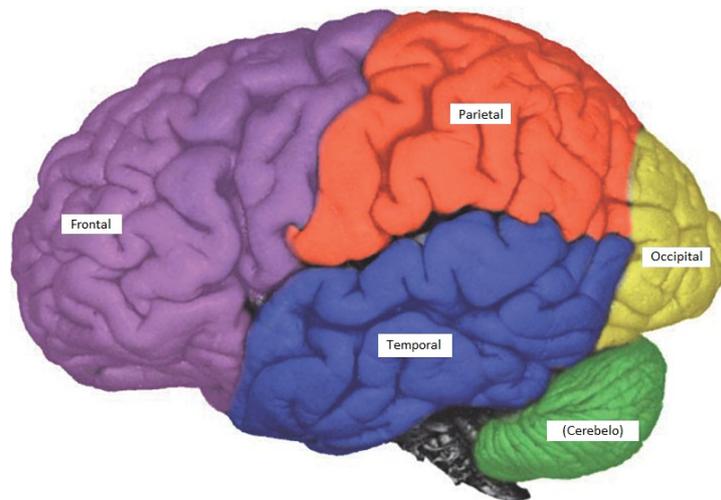


Figura 3.1. Encéfalo humano visto desde una perspectiva lateral. Se destacan en diferentes colores los cuatro lóbulos que componen el córtex cerebral, así como el cerebelo. Adaptado de Squire *et al.* (2008).

La capa más externa del cerebro recibe el nombre de corteza o córtex cerebral. Se trata de una capa compuesta por un gran número de circunvoluciones, de manera que sus neuronas pueden participar de un gigantesco número de sinapsis. El córtex puede ser subdividido en cuatro regiones conocidas como lóbulos frontal, parietal, occipital y temporal (figura 3.1.). Tal y como indican un gran número de estudios de los que enseguida hablaremos, existe un área del lóbulo frontal, muy desarrollada en los primates (en especial en el ser humano), que parece desempeñar un papel particularmente decisivo en los procesos volitivos. Se trata del córtex prefrontal (figura 3.2.). Otras áreas destacables del lóbulo frontal son el córtex motor primario y el córtex motor secundario. Este último incluye al área motora suplementaria.

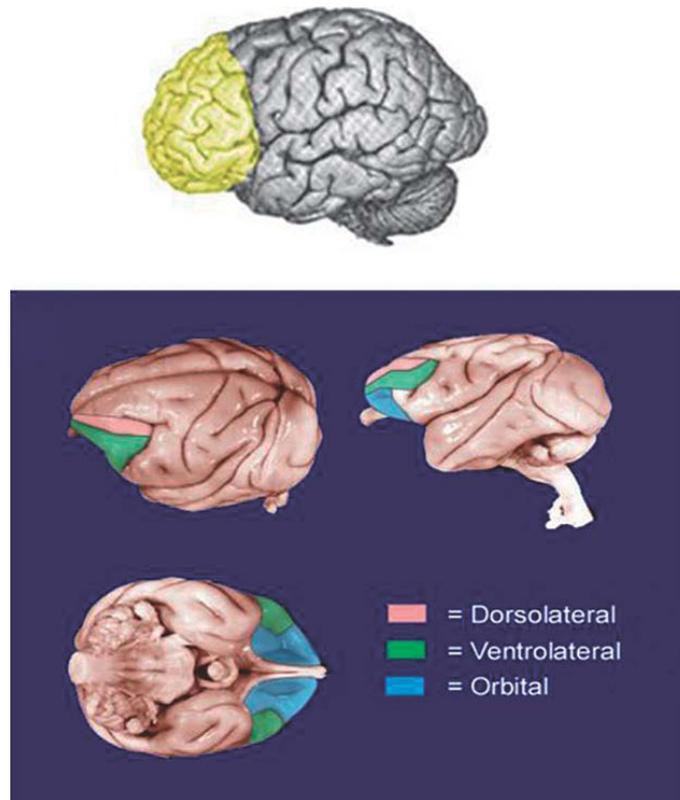


Figura 3.2. Arriba: cerebro humano en el que se destaca (en amarillo) la ubicación del córtex prefrontal. Abajo: cerebro de macaco, visto desde diferentes perspectivas, en el cual se muestran las principales zonas en que se divide su córtex prefrontal. Adaptado de Squire *et al.* (2008).

Existe una conexión muy importante entre el córtex cerebral en su conjunto y los ganglios basales, unas estructuras situadas en la zona subcortical que participan en la coordinación motora y en el aprendizaje. La información llega a los ganglios basales, procedente de la corteza, a través de uno de ellos principalmente: el cuerpo estriado (figura 3.3.). Este puede dividirse en dos partes: el núcleo lenticular y el núcleo caudado, que conecta con la amígdala. El córtex prefrontal, por su parte, guarda una serie de conexiones de distinta índole (Miller y Wallis 2008, pp. 1201-3) con sistemas de tipo sensorial y motor, con el sistema límbico (involucrado en las respuestas emocionales y formado por estructuras como el hipotálamo, el hipocampo y la amígdala,

entre otras) y con los ganglios basales (figura 3.3.). Esta variedad de interconexiones convierte al córtex prefrontal en un sistema multimodal y se considera un requisito indispensable para ejercer un control de la cognición (*ibid.*, pp. 1211-12). A la multimodalidad (multiplicidad de interconexiones) se unen propiedades de sus neuronas tales como la capacidad para mantener su actividad y conservar información, el procesamiento de reglas relacionadas con tareas, la flexibilidad para modificar su actividad en respuesta a cambios en las necesidades, o la codificación de información relacionada con recompensas. Todo ello hace que el córtex prefrontal sea una estructura fundamental para ejercer el control de la cognición y desempeñar un papel central en el procesamiento de comportamientos complejos (*ibid.*, pp. 1211-17).

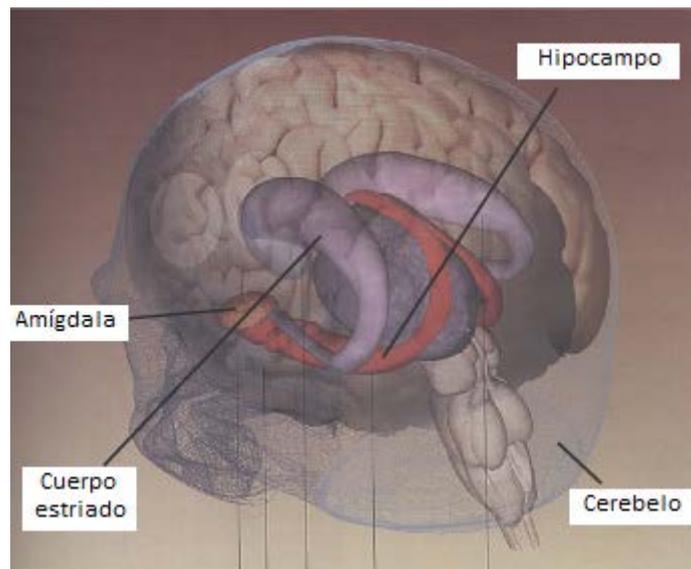


Figura 3.3. Ubicación del cerebelo y de algunas estructuras subcorticales del cerebro. Adaptado de Squire *et al.* (2008).

Dada la estrecha relación entre cognición y volición, y puesto que esta última da como resultado en numerosas ocasiones el desempeño de comportamientos complejos, los mencionados rasgos del córtex prefrontal lo convierten en una

estructura nerviosa a tener muy en cuenta en nuestro estudio. Es más, una abrumadora cantidad de estudios parece arrojar datos reveladores acerca de los papeles que el córtex frontal en general y el córtex prefrontal en particular (de ahora en adelante etiquetados como CF y CPF, respectivamente) juegan en los procesos volitivos, así como de los roles de otras estructuras relacionadas con ellos. Estos trabajos pueden ser clasificados según las técnicas utilizadas, destacando tres grandes grupos: estudios de lesiones, estudios de neuroimagen funcional y estudios de neuroelectrofisiología. Pero antes de proceder a detallarlos, es necesario realizar una aclaración sobre ellos: lo que muestran es que la actividad de ciertas áreas nerviosas es una *condición necesaria* para el desempeño de la volición, pero en ningún caso demuestran que dicha actividad sea una *condición suficiente* para esta tarea. En efecto, la correlación entre dos fenómenos (en este caso, la actividad nerviosa y los procesos volitivos) no indica por sí misma que uno de ellos sea la única causa necesaria para que el otro tenga lugar. Por ello, sería inconveniente posicionarse a favor de ciertas posiciones en el problema mente-cuerpo, como el materialismo eliminativo o la teoría de la identidad, basándonos *exclusivamente* en la correlación actividad nerviosa – volición. Una vez hecha la aclaración, sí es momento ya de hablar de los estudios de lesiones, de neuroimagen funcional y de neuroelectrofisiología.

### 3.3.2. Estudios de lesiones

En el presente apartado presentaremos estudios de casos en los que lesiones en el CF provocan efectos que nos dan información acerca de las funciones de este.

Damasio y van Hoesen (1983) presentan el caso de una mujer que, tras sufrir una lesión que afectó al cíngulo anterior (que forma parte de una circunvolución cerebral conocida como circunvolución del cíngulo, estrechamente relacionada con el sistema límbico, y que puede ser considerado

como una parte interna del CF), padeció un trastorno conocido como “mutismo acinético”. Las personas que sufren dicho trastorno son incapaces de moverse y comunicarse de forma normal a pesar de estar conscientes. La mujer del estudio reposaba en la cama y observaba a la gente que le hablaba sin poder contestar. Lo máximo que podía hacer era repetir a duras penas palabras o frases. Tras recuperarse de su lesión, la mujer reconoció que comprendía todo lo que le decían pero que era incapaz de aportar nada a la conversación, como si su mente estuviera en blanco. Esta mujer carecía de interés alguno para hablar o para reaccionar de alguna manera a lo que se le comunicaba, a pesar de que, en principio, era físicamente capaz de hacerlo.

Una patología asociada también a lesiones en el cíngulo anterior, así como en sus alrededores, es el síndrome de la mano extraña (o síndrome de la mano ajena) (Walter 2002, p. 571), popularizado por la película *¿Teléfono rojo? Volamos hacia Moscú*, dirigida por Stanley Kubrick. Este síndrome ha sido descrito por autores como Goldberg y Bloom (1990) o Gasquoin (1993). Los afectados mueven una de sus manos en contra de su voluntad, como si esta fuera de otra persona o tuviera vida propia. Otro síntoma típico consiste en que la mano afectada retiene un objeto con fuerza, pudiendo llegar a un punto en que el paciente sólo puede liberar dicho objeto utilizando la otra mano.

Las lesiones en el CPF pueden provocar diversos síntomas. Miller y Wallis (2008, pp. 1203-7) hablan de desinhibición en la conducta, cambios en el humor, incapacidad emocional, dificultades para planificar y organizar, y una inhábil memoria de trabajo (una memoria a corto plazo que se utiliza para situaciones tales como recordar una dirección de correo electrónico durante los instantes que se requieren hasta escribirla). También hablan del denominado “síndrome disejecutivo”, en el cual el paciente carece de control sobre sus funciones ejecutivas, entre las que se incluyen el comportamiento social y el comportamiento dirigido a objetivos. Un conocido caso, revisitado en profundidad por Damasio *et al.* (1994), fue el de un capataz de construcción ferroviaria llamado Phineas Gage, quien en un accidente laboral sufrió una

sería lesión en su CPF. Gage sobrevivió al accidente y, en muchos sentidos, su comportamiento siguió siendo perfectamente normal: tenía la misma inteligencia que antes del accidente, no había perdido la memoria, podía moverse y hablar sin ningún problema, adquirir nuevos conocimientos... Sin embargo, se convirtió en una persona soez, irreverente e incapaz de realizar tareas que requirieran planificación y organización, cuando antes del accidente se le tenía por una persona responsable y bien adaptada a las convenciones sociales.

Lhermitte (1983) describe cinco casos diferentes de lesiones en las áreas laterales del lóbulo frontal. Todos ellos guardan en común la manifestación de la denominada “conducta de utilización”, en la cual el afectado utiliza automáticamente un objeto por el simple hecho de serle presentado mediante estímulos visuales y/o táctiles. Lhermitte sugiere que estas lesiones tienen el efecto de eliminar la inhibición que el lóbulo frontal es capaz de ejercer sobre el parietal, desencadenando la acción de este último y haciendo que el paciente dependa enormemente de los mencionados estímulos.

Además de los hallazgos ya mencionados, los estudios de lesiones han contribuido a describir el fuerte vínculo existente entre las emociones y las decisiones de carácter moral. El trabajo de Koenigs *et al.* (2007) es un buen ejemplo de ello. Los autores hallaron que los individuos con lesiones en su CPF ventromedial, región que se considera fundamental para la aparición de emociones de tipo social, manifestaban una alta tendencia a esgrimir opiniones de tipo utilitario en situaciones morales con fuerte contenido emocional, mientras que sus opiniones en situaciones menos potentes emocionalmente se ajustaban más a la tendencia habitual.

### 3.3.3. Estudios de neuroimagen funcional

Las técnicas de neuroimagen funcional nos dan la oportunidad de observar qué cambios ocurren en el cerebro mientras tiene lugar una actividad, permitiéndonos detectar qué áreas cerebrales manifiestan activación y por tanto participan en ella. Es decir, nos proporcionan un seguimiento de la actividad cerebral.<sup>5</sup> Entre estas técnicas destacan especialmente dos: la tomografía por emisión de positrones, o PET (*Positron Emission Tomography*), y la imagen por resonancia magnética funcional, o fMRI (*functional Magnetic Resonance Imaging*). En la PET se inyecta por vía intravenosa un radiofármaco del cual se detecta su distribución en el cerebro antes de que sea metabolizado por el organismo.<sup>6</sup> En la fMRI no se inyecta ninguna sustancia, sino que se detectan cambios en el magnetismo provocados por el flujo sanguíneo en dirección a las áreas del cerebro que necesitan renovar su oxígeno porque presentan mayor actividad en ese momento.

Frith *et al.* (1991) realizaron un estudio mediante PET en individuos normales (sin lesiones cerebrales) mientras efectuaban misiones tales como producir una palabra o mover un dedo. Dicho estudio partía de la comparación entre actos rutinarios (como por ejemplo repetir palabras ya dadas) y actos volitivos (como pensar palabras partiendo solamente de su letra inicial). Lo que se encontró fue una activación, manifestada como un aumento del flujo sanguíneo, del CPF dorsolateral izquierdo durante los actos volitivos. Un patrón de activación similar ha sido hallado también en estudios posteriores mediante fMRI (Hyder *et al.* 1997). No obstante, tal y como señalan Spence y Frith (1999, p. 17), cabe preguntarse si la activación descrita por Frith *et al.*

---

<sup>5</sup> Para una introducción a la utilidad de los estudios de neuroimagen funcional, puede consultarse la obra *Pictures of the Mind*, de Boleyn-Fitzgerald (2010). Además, en ella aparecen varios de los estudios mencionados también en este apartado.

<sup>6</sup> Una técnica de neuroimagen funcional similar a la PET es la tomografía computerizada por emisión de fotones individuales o SPECT (*Single Photon Emission Computed Tomography*).

podría deberse en realidad a un mayor uso de la memoria de trabajo en comparación con los actos rutinarios. Por ejemplo, si el individuo ya produjo una palabra que empezaba por la letra “S”, podría haber intentado no utilizar la misma palabra si se le volvía a proponer la misma letra, y para ello se haría muy necesaria la memoria de trabajo, al contrario que para repetir palabras de manera rutinaria. Los sujetos de estudio, además, tenían que ceñirse a un plazo temporal para realizar las tareas encomendadas, y esto supone una restricción para el estudio, como señalan de nuevo Spence y Frith (*ibid.*, p. 19). Sin embargo, Jahanshahi *et al.* (1995) han demostrado que la activación del CPF dorsolateral se produce igualmente si el individuo es el encargado de elegir el momento de ejecución de un acto volitivo. Y hay más trabajos que delatan el papel del CPF dorsolateral en la volición. Dolan *et al.* (1993), por ejemplo, presentan un estudio realizado con treinta pacientes de esquizofrenia y cuarenta de depresión. Mediante PET detectaron cómo se reducía la actividad en el CPF dorsolateral izquierdo de los sujetos con pobreza del habla (dificultad para hablar con espontaneidad).

Un síntoma común entre los esquizofrénicos es la convicción por parte del afectado de que alguien externo a él posee el control de sus pensamientos y sus acciones. A esto se le llama “delirios de control ajeno”. Spence *et al.* (1997) rastrearon mediante PET, en dos ocasiones distintas y separadas entre sí unas semanas, la actividad cerebral de pacientes esquizofrénicos con delirios de control ajeno mientras estos manejaban un *joystick*. Además, la compararon con la de otros esquizofrénicos que no manifestaban dichos delirios, con la de individuos normales e incluso con la de ellos mismos en algunos casos en los que había tenido lugar una recuperación cuando se comenzó el estudio en la segunda de las mencionadas ocasiones. En los sujetos con delirios de control ajeno hallaron una especial actividad tanto en el córtex parietal como en el córtex del cíngulo. En el caso del cíngulo, ya se indicó en el apartado 3.3.2. su relación íntima con el CF. En cuanto al lóbulo parietal, “es un área de asociación heteromodal [...] involucrada en el procesamiento de información

sobre el cuerpo, y la programación de movimientos, en el espacio”, y por esta razón una actividad en ella que esté por encima de lo normal podría provocar cambios significativos en la intención de realizar un movimiento (Spence y Frith 1999, p. 25).

Los estudios de neuroimagen funcional, al igual que los de lesiones (tratados en el apartado 3.3.2.), han ayudado a dejar patente que, si bien la razón participa en los juicios morales, también existe una fuerte relación entre estos y las emociones. Sirva como muestra el trabajo de Greene *et al.* (2001). Se enfrentó a un grupo de estudiantes a una serie de problemas diversos que, según el grado de implicación emocional, fueron clasificados en tres clases: “no morales” (como escoger un tren o un autobús para viajar teniendo en cuenta algunas condiciones temporales), “morales impersonales” (como votar por el candidato político del cual se predice que será el responsable de un mayor número de muertes) y “morales personales” (como robar los órganos de alguien y repartirlos entre cinco personas). Para cada uno de los problemas se proponía una acción concreta y los sujetos de estudio debían responder si consideraban dicha acción “apropiada” o “inapropiada”. El mapeo mediante fMRI mostró que en los problemas de tipo “moral personal” se producía una activación de áreas del cerebro relacionadas con las emociones. En cambio, áreas importantes en la utilización de la memoria de trabajo manifestaban una menor activación que en los problemas de tipo “moral impersonal” y “no moral”.

#### 3.3.4. Estudios de neuroelectrofisiología

La electrofisiología es la disciplina que se ocupa de investigar la actividad eléctrica de las células y los tejidos del organismo. Cuando la investigación se realiza en el sistema nervioso, recibe el nombre de neuroelectrofisiología. Los estudios de esta especialidad han contribuido al avance en el conocimiento

acerca de los mecanismos nerviosos que subyacen al proceso volitivo. Esto queda patente en obras como la de Schultz (1999), donde el autor recapitula una serie de estudios realizados en su laboratorio. En dichos estudios se efectuó la medición de la actividad eléctrica de neuronas individuales situadas en el CF y en el cuerpo estriado del macaco cangrejero (*Macaca fascicularis*). De forma simultánea a la medición se hizo un seguimiento de los movimientos oculares y la actividad muscular. Todo ello con el objetivo de registrar lo que sucedía en comportamientos dirigidos a objetivos. Se pudo observar que, no habiendo estímulos externos que pudieran desencadenar la iniciación del comportamiento, tenía lugar una activación de las neuronas del cuerpo estriado unos segundos antes de que los primates iniciaran sus movimientos conductuales. Se detectó, además, una actividad parecida en el área motora suplementaria, la cual está íntimamente conectada al cuerpo estriado y junto con el conjunto de los ganglios basales forma un bucle de regulación de los movimientos. Por otra parte, cuando los animales detectaban una recompensa o esperaban recibirla se activaban también las neuronas del cuerpo estriado. Este hecho, según Schultz, podría deberse a que los primates efectuarían un proceso de evaluación de los posibles resultados previamente a la iniciación de su reacción comportamental. Si atendemos a todas estas observaciones, parece que el cuerpo estriado, puente de información entre los ganglios basales y el córtex cerebral, cumple en los primates un importante papel en el procesamiento de movimientos voluntarios y en la regulación del comportamiento dirigido a objetivos.

Hay más autores que han desarrollado su investigación en torno a la neuroelectrofisiología de la volición. Entre todos ellos destaca especialmente Benjamin Libet, cuyos trabajos no se centran en las estructuras implicadas en el proceso volitivo sino más bien en su dimensión temporal. La gran repercusión de estos trabajos justifica que les dediquemos una sección independiente más adelante (sección 3.4.).

### 3.3.5. Algunos comentarios adicionales

A tenor de la gran cantidad de trabajos expuestos a lo largo de toda la sección, quizá sea conveniente efectuar una pequeña recapitulación acompañada de algunas puntualizaciones necesarias para evitar caer en ciertos dogmatismos.

Los estudios de lesiones, de neuroimagen funcional y de neuroelectrofisiología que aquí se han expuesto dejan poco lugar a la duda: parece existir un vínculo innegable entre la actividad de ciertas áreas nerviosas y la génesis de decisiones de carácter voluntario. Entre todas estas áreas, parece que el CF juega un papel central en conjunción con otras estructuras estrechamente relacionadas con él, como el lóbulo parietal, la circunvolución del cíngulo (en especial el cíngulo anterior), los ganglios basales (entre los que destaca el cuerpo estriado) y la amígdala. Tal y como explica Walter (2001, p. 253), a grandes rasgos “la función del córtex frontal es *organizar el comportamiento a lo largo del tiempo*. Mientras que el área motora se preocupa de organizar y ejecutar movimientos, es trabajo del córtex prefrontal “controlar” procesos cognitivos que garanticen que movimientos apropiados son seleccionados en el momento correcto para el lugar correcto”. El CPF desempeña un rol especialmente importante, porque, en palabras de Spence y Frith (1999, p. 27), “parece estar involucrado en mantener posibles acciones en mente antes de que sean ejecutadas, y en seleccionar cuál será llevada a cabo” (si bien estos autores se refieren en concreto a la región dorsolateral del CPF). Es posible que la selección tenga lugar mediante la inhibición de todas las acciones posibles a excepción de una, que será la que finalmente se produzca (Goldman-Rakic 1987). Esta posibilidad estaría en sintonía “con la observación de que los pacientes con daño en el córtex prefrontal son a menudo incapaces de inhibir respuestas inapropiadas para su ambiente” (Spence y Frith 1999, p. 19), tal y como atestiguan casos como el de Phineas Gage, analizado a fondo por Damasio *et al.* (1994) y mencionado en el apartado 3.3.2. También podría ser compatible con la opción, propuesta por Benjamin Libet y tratada en la

siguiente sección, de que el inconsciente constituya una fuente de opciones distintas y nuestra voluntad consciente efectúe una selección en forma de veto.

Si bien es indudable la conexión entre la actividad de las áreas mencionadas y el desarrollo del proceso de volición, hay matices que sugieren evitar una excesiva simplificación de la situación:

1) Existe una inseparable vinculación entre la volición y la emoción, tal y como ponen de relieve dos de las publicaciones ya mencionadas en los apartados 3.3.2. (Koenigs *et al.* 2007) y 3.3.3. (Greene *et al.* 2001).<sup>7</sup> Además, parece cada vez más claro que “la memoria y la emoción están íntimamente conectadas en el cerebro, y alterar nuestra experiencia de una puede afectar drásticamente a nuestra experiencia de la otra” (Boleyn-Fitzgerald 2010, p. 90). Teniendo en cuenta esto, podemos pensar en una especie de triángulo sistémico que conecta de modo inextricable la memoria, la emoción y la volición.

2) Encontramos, según Walter (2001, p. 256), un problema de “regresión infinita” hacia los precursores de la voluntad en el cerebro. Fuster (1995, p. 296) nos avisa sobre ello de esta forma: “[A]signar la voluntad a cualquier región frontal del cerebro nos lleva obviamente a preguntarnos por una orden previa sobre esa región desde otra estructura; la misma cuestión puede ser planteada sobre esa otra estructura, cualquiera que esta fuera, y entonces sobre su precursora, y así sucesivamente”. El autor plantea atender a un ciclo que él bautiza como “ciclo percepción-acción”, en el cual existe un bucle de influencia mutua entre los córtex frontal y posterior y las estructuras subcorticales del cerebro, de manera que no podemos pensar en un “verdadero origen” (*id.*).

3) La propia metodología de los estudios podría ser responsable de que extraigamos conclusiones que nos conducen a localizar determinados procesos neurales en regiones concretas del CPF (Walter 2011, p. 524). De hecho, hay

---

<sup>7</sup> Sobre este asunto cabe destacar también la hipótesis de los marcadores somáticos, propuesta por Damasio (1994). Esta hipótesis será tratada en el apartado 4.1.3. del próximo capítulo.

teorías que presentan la actividad del CPF de un modo integrante (*e.g.*, Miller y Cohen 2001).

4) La fisiología y la parasitología nos proporcionan importantes datos que guardan relación con la volición. Parece ser que ciertas moléculas como la dopamina o la serotonina, entre otras, ejercen funciones de modulación en el sistema nervioso influyendo en procesos como la atención, la motivación, la impulsividad, etc. (Fellous 1999; Walter 2011, pp. 525-6). Además, la infección por el parásito *Toxoplasma gondii* repercute en factores como el nivel de dopamina, el tiempo de reacción e, incluso, las posiciones morales (Flegr *et al.* 1996; Flegr 2007, 2013).

Teniendo en cuenta los matices expuestos, podemos concluir que resultaría un tanto precipitado situar la volición, el libre albedrío o la conciencia en compartimentos estancos de nuestro sistema nervioso (cf. Boleyn-Fitzgerald 2010, p. 113). En lugar de esto, probablemente sería más prudente pensar que, si bien ciertas áreas resultan decisivas en la generación de nuestras decisiones voluntarias (en especial el CPF), dichas áreas no podrían ejercer esta función sin estar integradas en un amplio sistema formado por diversas estructuras, en el cual la cognición, la emoción y la memoria se encuentran fuertemente interrelacionadas. La capacidad de un área nerviosa para funcionar de forma independiente y para, a su vez, integrarse de forma coordinada en un conjunto es conocida como “metaestabilidad” (Kelso y Tognoli 2009). El mencionado sistema, además, no puede ser aislado completamente del individuo en su conjunto, sujeto a su vez a influencias del entorno en el cual se desenvuelve (la infección por *Toxoplasma gondii* sería un ejemplo de ello, aunque también debemos pensar en factores de tipo social, cultural, etc.). Por lo tanto, la actividad de las áreas nerviosas mencionadas es una *condición necesaria* pero no forzosamente *suficiente* para que tenga lugar el proceso volitivo. También hay que tener en cuenta que la correlación entre dos fenómenos cualesquiera (en nuestro caso, se trata de la actividad nerviosa y de los procesos volitivos) no

indica por sí sola que uno de ellos sea la única causa necesaria para que aparezca el otro (véase sección 3.3.1.).

### 3.4. Benjamin Libet y la génesis temporal de actos voluntarios

El interés de Libet en la electrofisiología de la volición se centra en la relación existente entre la aparición de un cambio eléctrico cerebral denominado “potencial preparatorio” (*readiness potential*, de ahora en adelante RP) <sup>8</sup> y el origen de los actos voluntarios. En Libet (1999) encontramos un análisis integrador de su investigación que, si bien menciona un gran número de trabajos previos (Libet *et al.* 1979, Libet *et al.* 1983, Libet *et al.* 1991; Libet, Wright y Gleason 1982, 1983; Libet 1985, 1989, 1993, 1994, 1996), nos será útil aquí como guía para exponer sus aportaciones de manera global.<sup>9</sup>

La definición de libre albedrío con la que Libet trabajó requería dos condiciones (1999, p. 47): (1) que el acto fuera endógeno en el sentido de que no hubiera constricciones que pudieran afectar al desempeño del acto voluntario, y (2) que el sujeto de estudio sintiera que poseía el control sobre dicho acto. En la mayoría de las pruebas no había restricción temporal, así que “los sujetos efectuaban un simple golpe o flexión de la muñeca [o los dedos] en cualquier momento que sentían el impulso o deseo de hacerlo” (*ibid.*, p. 49). El RP aparecía con anterioridad a la actividad de los músculos que participaban en la ejecución del acto (*ibid.*, pp. 48-9), pero tan importante como registrar dicha aparición era averiguar en qué momento eran los individuos del experimento conscientes de su decisión. Para ello se utilizó un reloj modificado en el cual un

---

<sup>8</sup> El RP es detectable en el cuero cabelludo y fue descrito por primera vez por Kornhuber y Deecke (1965), refiriéndose a él con el término alemán *Bereitschaftspotential*.

<sup>9</sup> Resulta también de gran interés su obra *Mind Time: The Temporal Factor in Consciousness* (Libet 2004).

punto de luz efectuaba un giro completo cada 2.56 segundos, y no cada 60 segundos como sucede en los relojes normales, con el resultado de que cada segundo imaginario del reloj adaptado se correspondía con 43 milisegundos reales (*ibid.*, pp. 49-50). Tras completar la misión de flexionar la muñeca o los dedos en el instante deseado, cada sujeto tenía que comunicar en qué posición del reloj modificado se encontraba el punto de luz en el momento en que era consciente de su deseo de efectuar dicha flexión (momento etiquetado por Libet con la letra W). Asimismo, la precisión con la que cada uno realizaba la mencionada comunicación fue debidamente calibrada, encontrando un error medio razonablemente escaso consistente en -50 milisegundos (*id.*). Todo el procedimiento de registro de los RP y de los tiempos en el reloj modificado fue realizado cuarenta veces para cada individuo, y se obtuvo un promedio de los resultados.

Libet descubrió dos tipos de RP (*ibid.*, p. 48). Por un lado estaban los que aparecían cuando el sujeto afirmaba haber tenido un plan para efectuar el movimiento dentro del plazo de un segundo (RP I). Por otro, aquellos que tenían lugar de forma espontánea a juicio del sujeto y, en consecuencia, sin haber efectuado ningún plan previo para ejecutar la acción (RP II). Los experimentos mostraron (*ibid.*, pp. 50-51) que en las flexiones espontáneas el RP II aparecía a los -550 milisegundos, previamente a la actividad de los músculos involucrados en el movimiento (considerada como tiempo 0). En el caso del RP I, registrado en las flexiones planificadas, el registro tenía lugar a los -1050 milisegundos. Además, se detectó la aparición de W a los -150 milisegundos (tras corregir su desviación de -50 segundos) independientemente de si la flexión era espontánea o planeada. Esquematizando:

RP I [-1050 ms] ----- W [-150 ms] --- Actividad muscular [0 ms]

RP II [-550ms] ----- W [-150 ms] --- Actividad muscular [0 ms]

Se observa una diferencia de 900 milisegundos entre RP I y W, y una de 400 milisegundos entre RP II y W. Esta última, apunta Libet, podría ser mayor porque “el proceso iniciador real en el cerebro probablemente comienza antes que nuestro RP registrado, en un área desconocida que entonces activa el área motora suplementaria en el córtex cerebral” (*ibid.*, p. 51). El RP, añade el autor, podría surgir precisamente del área motora suplementaria.

A la vista de los resultados obtenidos, Libet sostiene lo siguiente:

1) La iniciación de los actos voluntarios tiene lugar de manera inconsciente, ya que se produce antes de que el sujeto sea consciente de su decisión (*id.*).

2) La consciencia, a pesar de la conclusión anterior, podría intervenir en el proceso volitivo (*ibid.*, pp. 51-2). El sujeto dispone en principio de 150 milisegundos para intervenir de manera consciente. Este es el tiempo que media entre la aparición de W y la actividad muscular. No obstante, el córtex motor primario, mediante un proceso ya irreversible, se comunica con la región motora espinal durante los últimos 50 milisegundos previos a la actividad muscular. Por esta razón el plazo real queda reducido a 100 milisegundos. Durante este breve espacio de tiempo, el individuo podría vetar de forma consciente el proceso de movimiento muscular que había sido iniciado inconscientemente. Debido al diseño de los experimentos, en los que no se efectuaba registro eléctrico si no se realizaba un movimiento, no fue posible detectar el veto a pesar de que los participantes afirmaron en alguna ocasión haberlo llevado a cabo. Sí se pudo detectar, con otro tipo de diseño experimental, procesos de veto en movimientos previstos para ser realizados en un momento ya fijado. En estos casos sucedió que, tras la aparición de un RP prediciendo un movimiento, los sujetos vetaron este con un registro temporal que se encontró entre los 100 y los 200 milisegundos antes del mencionado momento.

3) No parece que la voluntad consciente pueda hacer las veces de “gatillo” que facilite que un proceso volitivo evolucione hasta el punto en el que tiene lugar una acción (*ibid.*, p. 52). Si se piensa en los actos voluntarios que realizamos de forma automática, se puede observar que pueden producirse sin que se detecte una intención preliminar consciente de ejecutarlos. Además, con anterioridad a este tipo de actos se detecta un RP que tiene una amplitud muy pequeña y una muy escasa duración (*id.*).

4) Cabe preguntarse si la aparición del veto puede tener un origen inconsciente, siendo el sujeto más tarde consciente de dicho veto, de un modo similar a lo que sucede con la actividad muscular estudiada (*ibid.*, pp. 52-3). Sin embargo, no sería necesario que esto sucediera. La razón es que *el veto constituye un mecanismo de control*, contrariamente a lo que sucede cuando se es consciente de querer hacer un movimiento, y “[n]o hay imperativo lógico en ninguna teoría mente-cerebro [...] que requiera actividad neural específica que preceda y determine la naturaleza de una función de control consciente” (*ibid.*, p. 53). Tampoco se ha demostrado experimentalmente que sean imprescindibles mecanismos inconscientes para que tenga lugar el mecanismo de control (*id.*). En cualquier caso, no podemos descartar que dichos mecanismos den lugar a factores sobre los cuales se fundamente la decisión de vetar y que sean necesarios para que el sujeto tenga conocimiento del veto que realiza, “pero el *contenido* de ese conocimiento (la decisión real de vetar) es una característica separada que no necesita tener el mismo requisito” (*id.*).

5) A pesar de que los actos investigados son eminentemente simples, los resultados tienen un gran impacto a la hora de considerar todos los tipos de actos voluntarios, independientemente de su grado de complejidad o planificación (*ibid.*, pp. 53-4). Distinguiendo la intención “de actuar ahora” [*sic*] de un eventual proceso de deliberación previa, y teniendo en cuenta que tanto el RP I (en casos de cierta planificación para actuar) como el RP II (en casos de acciones espontáneas) aparecen con anterioridad a W (deseo consciente de efectuar el movimiento), se puede decir que el orden temporal de sucesos

hallado para la intención “de actuar ahora” sería extrapolable a todos los actos voluntarios.

En conclusión, Libet considera que el libre albedrío tiene cabida en el ser humano, pero no como creador sino más bien como censor: los mecanismos inconscientes nos proporcionarían una fuente de opciones variadas, y entre ellas nuestra voluntad consciente, mediante el uso del veto, efectuaría la selección de cuáles terminarían finalmente desembocando en actos volitivos (*ibid.*, p. 54). Finalmente, el autor parece decantarse por una postura de tinte incompatibilista (*ibid.*, p. 56):

Mi conclusión sobre el libre albedrío, uno genuinamente libre en el sentido indeterminado, es entonces que su existencia es al menos una opción científica tan buena como lo es su rechazo por la teoría determinista, si no una mejor. Dada la naturaleza especulativa de las teorías tanto determinista como indeterminista, ¿por qué no adoptar la visión de que tenemos libre albedrío (hasta que alguna evidencia contradictoria real pueda aparecer, si alguna vez lo hace)? Tal visión nos permitiría al menos proceder de una manera que acepta y acomoda nuestra propia sensación profunda de que tenemos libre albedrío.

Los hallazgos de Libet han sido enriquecidos por otros autores mediante el análisis de sus experimentos y la búsqueda de un adecuado rigor conceptual que ayude a contextualizarlos. Gomes (1999, p. 65), por ejemplo, sostiene que el término “intención” puede dar lugar a diversas interpretaciones, y ofrece tres posibles significados:

- (1) “[L]a intención de actuar en algún momento en el futuro” (*id.*),
- (2) “la intención de actuar ahora (la cual podría todavía fracasar en conducir a una acción)”, y que no debemos confundir con “la decisión irrevocable de actuar ahora” (*id.*), o

- (3) “el objetivo o propósito de una acción” (*ibid.*, p. 66), con un sentido, por lo tanto, teleológico.<sup>10</sup>

El autor otorga relevancia al hecho de que en algunas de las pruebas diseñadas por Libet se hallara un RP precediendo a un veto en un movimiento concebido para su ejecución en un momento previsto, ya que demuestra que puede generarse un RP sin necesidad de que se termine efectuando un acto, y por esta razón no se debe considerar que el RP equivale a “la decisión irrevocable de actuar” (*ibid.*, p. 68).

También es importante para Gomes realizar una distinción entre los diferentes modos de interpretar los significados de “acto consciente”, “acto voluntario” y “decisión consciente” (*ibid.*, pp. 68-9). Así, el acto consciente podría ser:

- (1) “Un acto del cual somos conscientes” (*ibid.*, p. 68), o
- (2) “[u]n acto que es decidido tras la consciencia de la intención de actuar o tras imaginar conscientemente el acto” (*id.*).

El acto voluntario, por su parte, podría ser:

- (1) “Un acto que el sujeto siente que ha sido causado por una decisión consciente” (*ibid.*, p. 69), o
- (2) “[u]n acto que es decidido tras la consciencia de la intención de actuar o tras imaginar conscientemente el acto”, siendo lo mismo, por tanto, que un acto consciente entendido en su sentido (2) (*id.*).

En cuanto a la decisión consciente, podría ser:

- (1) “Una decisión de la cual nos volvemos conscientes, pero que podría existir y producir efectos antes de ello” (*id.*), o
- (2) “[u]na decisión que es intrínsecamente consciente, y en consecuencia sólo puede producir efectos después de que somos conscientes de

---

<sup>10</sup> Nótese que el autor se refiere en (1) a una intención previa (futura), en (2) a una intención inmediata (que puede ser de carácter irrevocable o no) y en (3) a una intención con la cual se pretende algo.

ella”, existiendo la posibilidad de que dicha decisión sea determinada o bien indeterminada (*id.*).

Gomes explica cómo, dependiendo del sentido que otorguemos a los términos “acto consciente”, “acto voluntario” y “decisión consciente”, puede ser muy distinta la manera en que concibamos las relaciones entre los fenómenos que denotan, dando lugar a diversas interpretaciones teóricas del proceso volitivo (*ibid.*, pp. 69-70). Asimismo, piensa que Libet tendría razón al afirmar que la iniciación de los actos voluntarios se produce inconscientemente si entendemos esta como la activación de señales que podrían desembocar en la acción (*ibid.*, p. 71), si bien podría considerarse consciente si tenemos en cuenta que el individuo es consciente del contexto en que está efectuando la acción (*ibid.*, pp. 72-3). El autor sostiene, finalmente, que no debemos confundir los actos conscientes con los actos volitivos (*ibid.*, p. 73) y que los únicos actos libres de verdad son los actos deliberados (*ibid.*, p. 75). Los actos volitivos no deliberados podrían tener, en palabras de Gomes, “un grado intermedio de libre albedrío”, pues en ellos, si bien gozamos de posibilidades alternativas <sup>11</sup> y somos conscientes del contexto en que actuamos, no tiene lugar una elección consciente de dichos actos (*id.*)

Mele, que es un firme defensor del papel causal fundamental que la voluntad consciente, en algunas ocasiones, juega en la ejecución de una acción subsiguiente (Mele 2009), ha sido especialmente fecundo en sus contribuciones al debate acerca del trabajo de Libet. Desde el punto de vista conceptual entiende “decisión práctica” como una acción mental que no se produce por deliberación y en la cual creamos una intención de actuar (Mele 2011, p. 501). Ahora bien, hay intenciones que no son fruto de una decisión, como sucede con algunas intenciones relacionadas con acciones que realizamos de forma

---

<sup>11</sup> Conviene recordar que hay posiciones compatibilistas en las cuales las posibilidades alternativas no son necesarias para decidir y/o actuar libremente. Los ejemplos basados en personajes, ejemplos de autocaptura y ejemplos estilo Frankfurt se han utilizado para argumentar en este sentido (véase sección 1.2.).

rutinaria (*id.*). Además, tomar la decisión de realizar una acción (la decisión práctica) no debe confundirse con querer llevarla a cabo, ya que en este segundo caso el vínculo con la acción final es más suave que en el primero (*id.*). También es importante para Mele efectuar una distinción entre las intenciones y decisiones de tipo “proximal” (relacionadas con acciones que se llevarán a cabo de forma inminente) y de tipo “distal” (relacionadas con acciones que se efectuarán más adelante) (*id.*).<sup>12</sup> Obsérvese que en los experimentos de Libet podríamos estar tratando con ejemplos del primer tipo, y así lo interpreta también Mele, quien cree que, aunque la existencia de intenciones inconscientes es objeto de discusión, las intenciones proximales de tipo inconsciente pueden tener lugar en algunos casos, y pone como ejemplo que conductores con experiencia pongan en marcha los intermitentes de un vehículo para avisar de que se va a efectuar un giro (*ibid.*, pp. 504-5). Sin embargo, reconoce que en los experimentos de Libet no puede descartarse una interpretación según la cual las activaciones musculares “no son producidas por intenciones proximales sino por la combinación de las intenciones condicionales de los sujetos de flexionar siempre que detectan un impulso proximal consciente de flexionar junto con su detección de tal impulso” (*ibid.*, p. 504). Según esta interpretación, el RP II, registrado a los -550 ms, podría estar “correlacionado con una causa potencial de un impulso proximal consciente de flexionar” (*id.*). Además, tal y como Mele recuerda, en los procesos de veto encontrados por Libet para movimientos que debían ser realizados en un momento determinado, los participantes de estas pruebas ya sabían que tenían que vetar el movimiento que estaba previsto para el momento acordado (*ibid.*, p. 506). Por esta razón, la aparición de un RP entre 100 y 200 segundos antes de dicho momento no tendría por qué corresponder a la intención de ejecutar el movimiento, pues sería posible que los individuos albergaran simultáneamente tanto dicha intención como la contraria, es decir,

---

<sup>12</sup> Nótese que el autor se refiere a intenciones/decisiones inmediatas y previas (futuras), respectivamente.

la de no realizar el movimiento (*id.*). El RP encontrado aquí no puede corresponder a una intención proximal de efectuar un movimiento, y sin embargo presenta unas características similares al RP I hallado en las pruebas en las que los participantes realizaban movimientos (*id.*). La conclusión, en opinión de Mele, es que cabe la posibilidad de que los RP I y RP II registrados en los casos de movimientos ejecutados tampoco correspondan a intenciones proximales de efectuar un movimiento, sino a sus causas potenciales (*ibid.*, pp. 506-7). Las intenciones proximales, en ambos casos, harían su aparición más adelante (*ibid.*, p. 507).

Hay estudios empíricos que parecen reforzar las opiniones de Mele. Haggard y Magno (1999) estudiaron los tiempos de reacción de siete personas al escuchar una orden. Si asumimos que la orden provoca la aparición de una intención proximal y que, además, se tardan unos milisegundos en que dicha orden sea debidamente recibida por el sujeto, observaremos que la intención proximal aparece en el entorno de los -200 ms (siendo 0 ms el momento en que tiene lugar la actividad muscular) (Mele 2011, p. 504). Esto dista bastante de los -550 ms a los que surge el RP II y, por el contrario, está muy próximo a los -200 ms (-150 ms, si incluimos la corrección hecha por Libet) a los que aparece W (momento en que el sujeto es consciente de su deseo de efectuar el movimiento) (*id.*). Trevena y Miller (2010, p. 447), por su parte, afirman que en su estudio “[n]o hubo evidencia de signos electrofisiológicos más fuertes antes de una decisión de mover que antes de una decisión de no mover, por lo que estos signos claramente no son específicos de la preparación del movimiento”.

Walter (2011, pp. 519-20) es muy crítico con el planteamiento metodológico de los experimentos de Libet. Considera que la intención consciente que tienen los participantes de efectuar un movimiento aparece en una fase inicial del experimento, ya que desde un primer momento conocen cuáles son las condiciones de este y tienen decidido realizar el movimiento posteriormente. Además, menciona la posibilidad de que el cambio eléctrico registrado sea en realidad la variación contingente negativa, o CNV (*contingent*

*negative variation*), y no el RP. La CNV es un cambio que tiene lugar entre un estímulo avisador y un estímulo imperativo. En las pruebas de Libet, el avisador podría aparecer en el momento en que el reloj modificado se ponía en marcha, mientras que el imperativo podría hacerlo en el momento en que el sujeto observa la posición del punto de luz cuando percibe el impulso de ejecutar el movimiento. Otro problema de los experimentos, a juicio de Walter, es que se produce la paradoja de que los participantes deben realizar movimientos espontáneos pero atienden a ciertas reglas, lo que hace que dichos movimientos sean simuladamente, y no realmente, aleatorios. Y no hay que olvidar, añade el autor, que los participantes podrían haber pensado que el impulso de ejecutar la acción debería tener lugar antes que esta, condicionando su percepción de la posición del punto de luz al aparecer dicho impulso. Es por todas estas razones que Walter escribe: “Problemas metodológicos así como de diseño muestran que es un fracaso la interpretación estándar de los experimentos tipo-Libet como un argumento en contra del papel de las intenciones conscientes y de este modo en contra de la realidad del libre albedrío” (*ibid.*, p. 520).

### **3.5. Daniel Wegner y la ilusión de la voluntad consciente**

En las discusiones recientes acerca del libre albedrío ha penetrado con fuerza la obra *The Illusion of Conscious Will* (Wegner 2002), que, dicho sea de paso, otorga una gran relevancia a los experimentos de Libet. Su autor, el psicólogo Daniel Wegner, sostiene en ella la idea de que las intenciones conscientes no juegan un rol causal en la ejecución de las acciones que las subsiguen. Ofrece argumentos que apoyan la idea de que los mecanismos que provocan que las acciones sean producidas por la mente se diferencian en anatomía y psicología de los mecanismos que conducen a que se produzca la voluntad consciente (*ibid.*, pp.

29-61), y afirma que esta última “es una ilusión en el sentido de que *la experiencia de querer conscientemente una acción no es una indicación directa de que el pensamiento consciente ha causado la acción*” (*ibid.*, p. 2). Tanto nuestra intención consciente de actuar como nuestra acción voluntaria tienen un origen inconsciente, pero cuando asociamos ambos causalmente, mediante un vínculo que realmente no es cierto, se produce la experiencia de voluntad consciente (*ibid.*, pp. 63-98). Esta ilusión está protegida por las personas “porque tienen un *ideal de agencia consciente* que guía sus inferencias sobre qué deben haber conocido y querido incluso cuando efectúan acciones que no pretendían” (*ibid.*, p. 146), así que “aspiran a ser agentes ideales que conocen todas sus acciones por adelantado” (*ibid.*, p. 145). Wegner, además, equipara la voluntad consciente con el libre albedrío (2004, p. 656): “[N]uestra discusión ha sido realmente *sobre* la experiencia de libre albedrío, examinando en detalle cuándo la gente lo siente y cuándo no. La idea especial que hemos estado explorando es explicar la experiencia de libre albedrío en términos de procesos deterministas o mecanicistas”.

Numerosas observaciones parecen jugar a favor de la postura de Wegner. Un ejemplo puede ser la conducta de utilización descrita por Lhermitte (1983) y mencionada en el apartado 3.3.2. Otro ejemplo podemos encontrarlo en el trabajo de Soon *et al.* (2008), quienes, mediante fMRI, observaron indicios de toma de decisiones, en forma de activación en el córtex prefrontal y parietal, con varios segundos de antelación al registro de la actividad consciente. Sin embargo, Wegner no está exento de críticos. Nahmias (2002, p. 536), por ejemplo, destaca la escasa importancia que este autor otorga a las intenciones conscientes vinculadas a acciones planeadas, deliberadas o prolongadas. También critica (*ibid.*, p. 533) que las “excepciones a la regla de que nuestras experiencias conscientes de nuestras acciones se ajustan a aquellas acciones” sean convertidas por Wegner en una norma general, ya que el hecho de que “la consciencia de querer una acción *pueda* ser separada de la acción no significa que cuando no están separadas, la consciencia no sea causalmente

relevante”. Otros autores han esgrimido razonamientos similares (*e.g.*, Pacherie 2006, p. 163). Hay trabajos, en cambio, que llegan a conclusiones similares a las de Wegner, pero con argumentos distintos a los suyos. Es el caso de Lau, Rogers y Passingham (2007), quienes, mediante un estudio realizado con una técnica estimuladora del córtex conocida como “técnica de estimulación magnética transcraneal”, o TMS (*Transcranial Magnetic Stimulation*), hallaron que el momento en que el sujeto experimentaba tener la intención de ejecutar una acción puede tener lugar con posterioridad a dicha acción. A juicio de los autores (*ibid.*, p. 81), esto podría apoyar la idea de que la voluntad consciente es una ilusión de un modo más convincente que el de Wegner. De todos modos, como ellos mismos reconocen, los resultados obtenidos no permiten descartar que el momento en que se experimenta la intención de actuar pueda estar presente también antes de la acción y desempeñar un papel causal en el proceso (*ibid.*, p. 89). Lo que sí permiten es contradecir la idea de que “la experiencia de intención [...] está completamente determinada antes de una acción” (*id.*). Hay que decir, sin embargo, que es controvertida la concepción de la intención como una experiencia. Tal y como destaca Moya (2012, p. 112), Wittgenstein sostenía que las intenciones (entendidas como propósitos) no son experiencias o sensaciones, ya que no poseen duración temporal.

### 3.6. Neurofilosofía y causalidad descendente <sup>13</sup>

Schwartz (1999) ha demostrado que es posible modificar la sintomatología y los circuitos cerebrales asociados al trastorno obsesivo-compulsivo entrenando la atención del afectado mediante terapia cognitivo-conductual. Otros estudios han puesto de manifiesto que, con un debido entrenamiento, se puede

---

<sup>13</sup> En esta sección efectuamos una aproximación empírica y breve a la causalidad descendente. Para un tratamiento pormenorizado de este fenómeno, véase el próximo capítulo.

controlar voluntariamente la activación del córtex del cíngulo anterior rostral hasta lograr una disminución en la percepción del dolor (deCharms *et al.* 2005). También hay pruebas de que, mediante estrategias cognitivas de regulación emocional, es posible modificar la manera en que se conecta la amígdala con el CPF y, así, disminuir el efecto diferencial que la genética provoca en el nivel de respuesta de la amígdala de distintos individuos a estímulos de carácter indeseable (Schardt *et al.* 2010). Los ejemplos mencionados no son más que una pequeña muestra de la creciente cantidad de estudios que atestiguan la capacidad del ser humano para modificar el funcionamiento de su sistema nervioso, denominada “neuroplasticidad”. Como destaca Boleyn-Fitzgerald (2010, p. xii):

Ahora podemos observar nuestros cerebros en pantalla, sanando y adaptándose a desafíos, y vemos que nuestros genes y experiencias tempranas absolutamente *sí* influyen nuestra conformación cognitiva y emocional de importantes maneras, y que absolutamente *no* llegan a dictar en quiénes nos convertimos. Esta previamente no apreciada flexibilidad y capacidad de entrenamiento de las vías neurales se denomina “neuroplasticidad”, y ha transformado la neurociencia moderna en un campo intensamente optimista donde los investigadores buscan nuevas técnicas diagnósticas y terapias para que los pacientes se recuperen del daño estructural y los desequilibrios químicos debidos a lesión cerebral traumática, apoplejía, enfermedad de Alzheimer, desórdenes emocionales, drogadicción, y dolor crónico. La investigación en neuroplasticidad no está revelando solamente cómo podemos curar lesiones cerebrales y agudizar nuestra mente, sino también cómo podemos fortalecer vías neurales claves para volvernos más felices, más amables, menos temerosos, y más eficaces simplemente cambiando la forma en que percibimos el mundo y nuestras reacciones a él.

Frith (2009) subraya el control que el contexto social y cultural ejerce en la voluntad del individuo, y considera que es en dicho contexto donde debemos enmarcar la interacción que se produce entre experimentadores y sujetos de estudio en los experimentos relativos a la ejecución de actos volitivos. Podríamos decir que dicha interacción es extensible a los tres ejemplos mencionados previamente, ya que las prácticas llevadas a cabo por los participantes difícilmente se pueden concebir sin tener en cuenta a los

experimentadores. En este sentido, podemos considerar que tiene lugar un influjo causal procedente del ambiente que rodea a dichos participantes, con el resultado de modificar mecanismos neurales. Por otra parte, el control del entorno queda patente también en trabajos que no son propiamente de neurociencia. Por ejemplo, Kurzban, DeScioli y O'Brien (2007) observaron que en sus experimentos el castigo de carácter moral era mayor si era infligido bajo observación. Vohs y Schooler (2008), por su parte, mostraron que el hecho de creer en el determinismo hace que sea más probable comportarse de forma fraudulenta. Si aceptamos que nuestro entorno sociocultural desempeña un papel causal importante en la adopción de nuestras creencias, podemos considerar que este estudio de filosofía experimental <sup>14</sup> encaja con la idea de que la cultura y la sociedad ejercen su influjo causal, vía creencias, en la voluntad del individuo. Por tanto, puede concebirse dentro de la intrincada red social y cultural el hecho de que las opiniones que tenemos en relación al libre albedrío parecen ser un factor que participa en la elección de nuestras acciones.

La realidad y las relaciones causales dentro de su seno pueden ser descritas en forma de jerarquía. Ellis (2009, p. 64) enumera ocho niveles en función de las disciplinas que los estudian: (1) física de partículas, (2) física atómica, (3) química, (4) bioquímica, (5) biología celular, (6) fisiología, (7) psicología, y (8) sociología, economía, política. Si pensamos en un universo organizado de forma jerárquica, podemos decir que la aproximación que, con la ayuda de la neurofilosofía mínima, habíamos realizado hasta ahora al problema del libre albedrío se había basado tanto en un sentido ascendente de la causalidad, desde nuestro cerebro hacia nuestro comportamiento, como en una causalidad dentro de un mismo nivel, que se manifiesta en la relación entre distintas regiones nerviosas (véase apartado 3.3.5.). Sin embargo, si tenemos en cuenta la neuroplasticidad y la influencia causal del entorno sociocultural sobre la voluntad del individuo apreciaremos un tipo de aproximación que afronta el

---

<sup>14</sup> Se pueden conocer más detalles acerca de la filosofía experimental en Knobe y Nichols (2011) y Nahmias (2011), por ejemplo.

problema desde una “causalidad hacia abajo” (*downward causation*), concepto introducido por Campbell (1974) y utilizado por muchos otros (*e.g.*: Kim 1992, Ellis 2009).

Si bien se trata de un concepto estelar entre los autores partidarios de la emergencia, la causalidad descendente no tiene por qué ser entendida solamente dentro de un ámbito emergentista.<sup>15</sup> Por otra parte, la organización jerárquica del universo puede servir como argumento para opinar en contra de la propia emergencia o, en general, en contra de la causalidad hacia abajo. Esto se puede observar en el trabajo de Moya (2011), quien ha analizado el reto que supone para la causalidad mental la concepción jerárquica según la cual hay una dependencia unidireccional de los niveles superiores respecto de los inferiores. Por lo tanto, podemos comprobar que la jerarquía del universo no tiene por qué ajustarse a una posición ontológica concreta sino a una descripción abarcable de la realidad: diferentes disciplinas del saber estudian realidades de distintas características. Pero si tenemos en cuenta los hallazgos mencionados anteriormente, que parecen mostrar la existencia de fenómenos de causalidad descendente, tendremos que reconocer que el reduccionismo no basta para afrontar la neurofilosofía del libre albedrío: no todo puede explicarse desde abajo hacia arriba. La causalidad en sentido ascendente es, no cabe duda, imprescindible para comprender la volición. Pero es solo una parte de la historia. Si pretendemos describir la causalidad en el fenómeno volitivo de una forma lo más completa posible, debemos adoptar una triple perspectiva: ascendente, intranivel y descendente (cf. Ellis 2009, pp. 78-9).

Es muy importante recalcar que el hecho de negar la suficiencia de la reducción en la neurofilosofía del libre albedrío no significa que necesariamente debamos descartar la verdad del determinismo. Un determinista podría alegar que la influencia del contexto cultural y social sobre la voluntad del individuo y sus vías neurales puede estar determinada desde un

---

<sup>15</sup> Basamos esta afirmación en la adopción del sentido tradicional del término “emergencia”. Para una visita más detallada a este concepto, véase el apartado 4.2.2.

origen externo al agente volitivo. Si bien es cierto que la reducción es empleada a menudo para defender el determinismo, no debemos caer en el error de confundir ambos conceptos o de enlazarlos de modo inseparable.

Nótese que la causalidad descendente no entra en conflicto con la aproximación que nos brinda la neurofilosofía mínima en el sentido de que no nos empuja a decantarnos por una posición concreta en el problema mente-cuerpo entre las que otorgan una gran importancia a las contribuciones de la neurociencia (véase sección 3.2.). Sin embargo, choca seriamente con el epifenomenismo y el materialismo eliminativo: el primero defiende un sentido exclusivamente ascendente de la causalidad (desde el cerebro hacia la mente, nunca viceversa), mientras que el segundo niega la existencia de los estados y procesos mentales (de la mente, en definitiva). El fenómeno de la neuroplasticidad, consistente en la capacidad para modificar el funcionamiento del sistema nervioso mediante el entrenamiento y el esfuerzo volitivo, parece jugar en contra de ambas posturas. Llegados a este punto, y si se acepta la causalidad descendente, vemos que, a pesar de utilizar la neurofilosofía mínima como inspiración inicial para no decantarnos por ninguna de las posiciones mente-cuerpo que toman en seria consideración los descubrimientos de la neurociencia, resulta inevitable efectuar una elección metafísica consistente en dejar a un lado dos de dichas posiciones.

### **3.7. Neurofilosofía y libre albedrío: análisis crítico**

Al principio del presente capítulo hemos presentado muy brevemente las posiciones más destacadas en relación al problema mente-cuerpo. A continuación hemos defendido la idea de Walter (2001, pp. 127-34) de que la neurofilosofía mínima es una herramienta útil como inspiración a la hora de aproximarnos desde una posición neurofilosófica al problema de la libertad sin

que sea necesario tomar partido por una de estas posiciones en concreto. Tras esto, hemos tratado las bases biológicas de la volición, centrándonos en los estudios de lesiones, de neuroimagen funcional y de neuroelectrofisiología. Hemos proseguido con la presentación de los experimentos de Benjamin Libet acerca de la génesis temporal de los actos voluntarios, completándola con opiniones que creemos relevantes en el debate relacionado con dichos experimentos. La posición de Daniel Wegner, defendiendo que la voluntad consciente es una ilusión, ha sido el siguiente tema tratado. Y después de ello, hemos reclamado la relevancia de la causalidad de tipo descendente a la hora de tratar la neurofilosofía del libre albedrío. Nuestra tarea durante la presente sección consistirá en utilizar toda esta información para analizar la relevancia que la neurociencia posee en el debate general sobre el libre albedrío y la responsabilidad moral (tratado en el capítulo 1). Esta tarea la efectuaremos a partir de las cinco preguntas de Kane acerca de la libertad: (A) La Cuestión de Compatibilidad, (B) La Cuestión de Significatividad, (C) La Cuestión de Inteligibilidad, (D) La Cuestión de Existencia, y (E) La Cuestión Determinista.

### *3.7.1. Neutralidad sobre la compatibilidad*

Kane (1996, p. 13) plantea el problema de la compatibilidad del libre albedrío de la siguiente forma:

[A] La Cuestión de Compatibilidad: ¿Es el libre albedrío compatible con el determinismo?

Diversos estudios de lesiones, de neuroimagen funcional y de neuroelectrofisiología han dejado patente la participación de ciertas áreas nerviosas en la función volitiva. Entre todas estas áreas nerviosas destaca particularmente el CPF. Si bien no debemos pasar por alto la dificultad que entraña situar anatómicamente el libre albedrío en compartimentos cerrados

del sistema nervioso ni la influencia causal de sentido descendente que nuestro entorno sociocultural ejerce sobre nuestra voluntad y nuestras vías neurales, parece obvio que sin el CPF y otras estructuras no sería biológicamente posible ejercer la volición: su actividad es una *condición necesaria* para que tenga lugar el proceso volitivo. No obstante, como ya vimos, no es obligatoriamente una *condición suficiente*. Por esta razón, podemos decir que no se puede responder “sí” o “no” a la Cuestión de Compatibilidad basándonos exclusivamente en la neurociencia: el posicionamiento a favor del compatibilismo o del incompatibilismo necesita de argumentos distintos y complementarios a los que los descubrimientos de la neurociencia pueden proporcionarnos. Pero imaginemos por un momento que la actividad del CPF y otras áreas sí fuera condición suficiente para la volición y que, además, los mecanismos biológicos que tienen lugar en esos lugares tuvieran un carácter determinista (algo que no se deduce de su mera actividad conducente a resultados volitivos). A simple vista, da la sensación de que estos hechos conllevarían que el libre albedrío no solamente sería compatible con el determinismo, sino que incluso sería imposible sin él: no podríamos ejercer nuestra libertad sin una determinación causal a nivel nervioso. Sin embargo, sucede lo contrario por dos razones. En primer lugar, un incompatibilista sostendría que una voluntad determinada por nuestras características biológicas no es una voluntad realmente libre, y que por lo tanto el libre albedrío y el determinismo no son compatibles. La postura de Daniel Wegner, sin ir más lejos, encaja en el marco incompatibilista (concretamente en el determinismo duro), ya que no duda en afirmar que nuestra voluntad está determinada por nuestro inconsciente y que nuestra voluntad consciente, que él equipara al libre albedrío, es una ilusión, aunque no hay que olvidar, como hemos dejado patente, que este autor ha recibido serias críticas. En segundo lugar, el determinismo de tipo neural, a pesar de lo que podría parecer, resulta amenazador para el compatibilismo, ya que este apuesta por una causalidad en la que nuestras acciones nacen de nuestros deseos, nuestros valores, nuestras decisiones, etc. (Moya 2012, p. 114). Por lo tanto, si

el determinismo neural fuera cierto (algo que, insisto, no se sostiene basándonos exclusivamente en la actividad neural conducente a resultados volitivos), la respuesta a la Cuestión de Compatibilidad sería “no”.

### 3.7.2. Relevancia parcial sobre la significatividad

El problema de la significatividad del libre albedrío es presentado así por Kane (1996, p. 13):

[B] La Cuestión de Significatividad: ¿Por qué queremos, o deberíamos querer, poseer un libre albedrío que es incompatible con el determinismo? ¿Es este un tipo de libertad “que valga la pena querer” (por utilizar la útil frase de Dennett)? Y si es así, ¿por qué?

Según Wegner, nuestra intención consciente de actuar y nuestra acción voluntaria se originan de modo inconsciente, pero al establecer un vínculo causal irreal entre ambos tiene lugar la experiencia de voluntad consciente. La voluntad consciente, por lo tanto, es una ilusión. Además, recordemos que este autor equipara la voluntad consciente al libre albedrío (Wegner 2004, p. 656): “[N]uestra discusión ha sido realmente *sobre* la experiencia de libre albedrío, examinando en detalle cuándo la gente lo siente y cuándo no. La idea especial que hemos estado explorando es explicar la experiencia de libre albedrío en términos de procesos deterministas o mecanicistas”. Lo que Wegner parece decir es que el libre albedrío entendido en un sentido indeterminista (es decir, el libre albedrío libertarista) es una ilusión. La Cuestión de Significatividad alude, en parte, a las razones por las cuales queremos disfrutar de una libertad de este tipo. La respuesta del propio Wegner gira en torno a la idea de que la mencionada ilusión se encuentra protegida por las personas “porque tienen un *ideal de agencia consciente* que guía sus inferencias sobre qué deben haber conocido y querido incluso cuando efectúan acciones que no pretendían” (*ibid.*, p. 146), de modo que “aspiran a ser agentes ideales que conocen todas sus

acciones por adelantado” (*ibid.*, p. 145). Es probable que dicho “ideal de agencia consciente” parta de la influencia del entorno sociocultural en el cual las personas se desenvuelven como agentes volitivos. En la sección 3.6. abordamos, de hecho, la influencia causal descendente desde dicho entorno hacia la voluntad y los mecanismos neurales individuales. He aquí una posible respuesta, fundamentada en hallazgos de la neurociencia, a una parte de la Cuestión de Significatividad. Pero es importante aclarar algo: el hecho de plantear que el trasfondo sociocultural puede tener una influencia causal en las razones que nos empujan a querer disfrutar de una libertad incompatible con el determinismo *es independiente del hecho de que este tipo de libertad exista o sea una ilusión*. Es una posibilidad, por qué no, que la influencia del contexto social y cultural esté complementada por la existencia real de una libre voluntad al estilo libertarista, o incluso surja de nuestras intuiciones colectivas acerca de esta. En definitiva, debe quedar claro que la respuesta a la Cuestión de Significatividad no depende de un posicionamiento acerca de la Cuestión de Existencia.

La Cuestión de Significatividad no sólo alude a las razones que nos empujan a querer una libertad incompatible con el determinismo. También pregunta si deberíamos querer, y por qué, dicha libertad. El libre albedrío incompatibilista ha sido concebido a menudo como valor originador de otros valores importantes para el ser humano (como la responsabilidad moral, la creatividad, la dignidad, etc.), aunque varios autores han negado este vínculo (Kane 1996. pp. 15-16). Sea como fuere, este componente de la Cuestión de Significatividad ya no puede ser tratado desde el prisma de la neurociencia, puesto que guarda relación más bien con la conveniencia en los niveles moral y social del posicionamiento a favor de un libre albedrío libertarista. De todos modos, y aunque se escapa del ámbito de la neurociencia, quizá sea interesante recordar el trabajo de filosofía experimental de Vohs y Schooler (2008), mencionado en la sección 3.6., quienes mostraron que el hecho de creer en el determinismo hace que sea más probable comportarse de forma fraudulenta.

Teniendo en cuenta esto, es una posibilidad a tener en cuenta que la creencia en una libertad incompatible con el determinismo sea un factor que influye de forma beneficiosa en nuestro comportamiento moral y social.<sup>16</sup>

### *3.7.3. Dudas sobre la inteligibilidad*

El contenido de la Cuestión de Inteligibilidad (Kane 1996, p. 13) es el siguiente:

[C] La Cuestión de Inteligibilidad: ¿Podemos darle sentido a una libertad o libre albedrío que es incompatible con el determinismo? ¿Es dicha libertad coherente o inteligible? ¿O es, como reivindican muchos críticos, esencialmente misteriosa y terminantemente oscura?

Esta cuestión puede resumirse preguntando: ¿Es inteligible una libertad incompatibilista? Los libertaristas son los únicos incompatibilistas que creen que los humanos tenemos libre albedrío, de forma que una pregunta casi equivalente sería: ¿Es inteligible un libre albedrío libertarista? Abrazar el libertarismo supone admitir que las posibilidades alternativas y el control último son necesarios para la libertad y que el determinismo excluye ambas condiciones (véase sección 1.3.). Un “sí” a la Cuestión de Inteligibilidad con la neurociencia como cimiento requiere explicar cómo es posible que un indeterminismo de origen nervioso haga posibles las posibilidades alternativas y el control último sobre nuestras decisiones y/o acciones. De lo que aquí se trata es de ofrecer una explicación que sea inteligible (concebible y coherente) y no de demostrar que se corresponde con la realidad. Por esta razón, no es necesario ser un libertarista para responder afirmativamente a la cuestión. Un

---

<sup>16</sup> En relación a este asunto, conviene recordar la tesis del ilusionismo propuesta por Smilansky (2000), basada en la creencia de que es necesario mantener, a escala social y moral, la ilusión de que el libre albedrío existe, pues sería nefasta para nosotros una creencia contraria extendida de modo generalizado (véase sección 1.3.).

determinista duro podría, por qué no, responder en ese mismo sentido, pero estaría obviamente en contra de la existencia de una libre voluntad incompatible con el determinismo.

Entre las explicaciones ofrecidas a lo largo del capítulo, una muy importante que podría ser utilizada por el libertarismo para reforzar su posición, aunque también ha sido utilizada por autores de otras tendencias, es la de Benjamin Libet. Dicho autor, que afirma que la iniciación de los actos voluntarios es inconsciente y anterior al momento en que somos conscientes de ellos, defiende también, no obstante, que dichos actos podrían ser vetados conscientemente. Piensa que el inconsciente podría constituir un manantial de opciones de acción entre las cuales nuestra voluntad consciente realizaría, como función de control, una selección en forma de veto. No obstante, podríamos dudar de la existencia de un número mayor que uno de opciones de acción en el seno de los experimentos de Libet, en los que, como destaca Walter, los participantes conocen desde un primer momento cuáles son las condiciones de estos y tienen decidido realizar el movimiento posteriormente. Además, si el veto se realiza de manera consciente, ¿cómo es posible que se pueda seleccionar conscientemente entre opciones de naturaleza inconsciente? Parece existir un vacío explicativo entre ambas fases, con dificultades de inteligibilidad y no aclarado por Libet. Por otro lado, nótese que, a simple vista, las opciones de acción podrían considerarse como posibilidades alternativas y la selección en forma de veto como control último (recordemos que las posibilidades alternativas y el control último son condiciones necesarias para la libertad según el libertarismo). Sin embargo, es una posibilidad que el mecanismo de veto tenga lugar sin que exista un control último, capacidad que consiste en ser la fuente, el autor, el origen último de las propias decisiones y/o acciones. Un libertarista consideraría que un veto consciente determinista no constituye en absoluto un control último, y en el caso de un veto consciente indeterminista tendría que lidiar con el argumento de *Mind*, según el cual no hay control último si la voluntad surge del azar. Es importante aclarar que

Libet no extrae de sus experimentos una vinculación del indeterminismo con la consciencia en nuestros actos (Libet 1999, p. 55), pero sí piensa que decantarse por un libre albedrío indeterminista es igualmente válido, incluso preferible por ciertas razones, que rechazarlo (*ibid.*, pp. 56-7). Por lo tanto, su opinión supone un guiño al libertarismo. No obstante, y dado que el propio Libet no lo hace, los libertaristas deberían encontrar una forma coherente de explicar de qué manera podría el indeterminismo intervenir en el proceso descrito por Libet si quisieran integrar dicho proceso en sus propias argumentaciones. Y la dificultad de esta intervención estriba en que debería ser explicada desde dos perspectivas: (1) cómo el *inconsciente* puede generar opciones de acción de forma indeterminista, y (2) cómo el proceso *consciente* de veto puede ser iniciado de forma indeterminista. Es decir, el indeterminismo tendría que estar en el origen de procesos tanto inconscientes como conscientes, además de sortear el argumento de *Mind*, para que la postura de Libet pudiera ser integrada por el libertarismo. Es conveniente añadir que, si diéramos por sentado (a pesar de las dificultades mencionadas anteriormente) que las opciones de acción son posibilidades alternativas y que el veto constituye un control último, y dado que los libertaristas rechazan la compatibilidad del determinismo tanto con unas como con otras, no bastaría con mostrar la coherencia de (1) o de (2): tendría que hacerse con ambas. Hacerlo solo con una de ellas quizá permita responder de forma afirmativa a la Cuestión de Inteligibilidad si entendemos esta de forma literal, ya que en ella se pregunta si tiene sentido una libertad incompatible con el determinismo. Parecería que, en principio, bastaría con dejar patente la coherencia de (1) o de (2) indistintamente. Pero teniendo en cuenta que la Cuestión de Inteligibilidad suele ser asociada con las posiciones libertaristas en la literatura acerca del libre albedrío, parece lógico pensar que responder con un “sí” a dicha cuestión debe partir de una demostración de coherencia de (1) y de (2) conjuntamente. Y esta es una tarea que, hasta la fecha, no se ha podido realizar. Sin duda, el argumento de *Mind* supone el más importante reto al respecto. No olvidemos, además, los problemas de

interpretación de los resultados de Libet, puestos de relieve por autores como Gomes, Mele o Walter. Podemos concluir, por tanto, que es muy problemático utilizar la postura de Libet, bien por sí sola o bien como apoyo adicional, para dar una respuesta afirmativa a la Cuestión de Inteligibilidad.

Podríamos preguntarnos si la influencia causal descendente desde el entorno social y cultural hacia la voluntad y las vías neurales individuales podría ser un refuerzo para el libertarismo. Para ello, los libertaristas deberían explicar de forma inteligible y coherente cómo es posible que el indeterminismo esté detrás de una influencia causal de nuestro entorno sociocultural, ya que dicha influencia podría ser considerada por muchos como determinismo social. No olvidemos, además, que el fenómeno mismo de la causalidad hacia abajo es un hecho discutido y, como observa Moya (2011), desafiado por concepciones jerárquicas en las que hay una dependencia unidireccional de los niveles superiores a partir de los inferiores. Y, por supuesto, una influencia causal indeterminista del entorno sociocultural no se libraría del desafío que constituye el argumento de *Mind*. Por último, es muy dudoso que un agente pueda gozar de control último si los mecanismos indeterministas que intervienen en el proceso volitivo son externos a él. En consecuencia, la causalidad descendente por sí misma no es suficiente para responder de forma afirmativa a la Cuestión de Inteligibilidad, y necesitaría una explicación complementaria fuertemente convincente y coherente.

#### 3.7.4. *Dificultades para aceptar la existencia*

Kane (1996, p. 13) formula la hipotética existencia de un libre albedrío libertarista del siguiente modo:

[D] La Cuestión de Existencia: ¿Dicha libertad existe en realidad en el orden natural? Y si es así, ¿dónde?

La Cuestión de Inteligibilidad tiene que ver con la coherencia de las explicaciones libertaristas, independientemente de la existencia o inexistencia de las libertades que proclaman. Este dilema queda reservado a la Cuestión de Existencia, que, contextualizada en el capítulo, podría formularse así: “¿Existe en el orden natural una libertad incompatible con el determinismo y compatible con la neurociencia?”. Ya que los libertaristas son los únicos que apuestan por la existencia de una libertad de este tipo, de lo que realmente se trata aquí es de explorar la existencia de un libre albedrío de tinte libertarista. Como vimos en el apartado anterior, la teoría de Libet acerca de la génesis temporal de los actos voluntarios podría ser utilizada por los libertaristas si consideran el mecanismo de veto como una manifestación de nuestra libertad y si demuestran que dicho mecanismo tiene un cimiento indeterminista. Dado que de lo que aquí se trata es de averiguar si dicha libertad existe o no, podemos decir que las averiguaciones de Libet acerca de cómo ejercemos el veto sobre algunas decisiones parecen ser un indicio a favor de dicha existencia. Sin embargo, y a raíz de las posturas mantenidas por autores como Gomes, Mele o Walter, que pudimos conocer en la sección 3.4., han quedado patentes los problemas para interpretar los experimentos de Libet. Y lo que es más importante: el propio Libet no se centra en averiguar si el veto consciente funciona de modo determinista o indeterminista (Libet 1999, p. 55). Es por ello que debemos concluir que no podemos responder positivamente a la Cuestión de Existencia acudiendo a los hallazgos de Libet: además de que presentan serios problemas de interpretación, no podemos dar por sentado que esconden procesos de naturaleza indeterminista.

### 3.7.5. Inconclusión sobre el determinismo

Kane (2002a, p. 6) se pregunta de la siguiente forma acerca de la verdad del determinismo como ley universal:

[E] La Cuestión Determinista: ¿Es el determinismo cierto?

Si gracias a la neurociencia obtuviéramos una respuesta negativa a esta cuestión, podríamos considerar que la solución es definitiva: si se encuentra un solo caso de indeterminismo, el determinismo ya no será cierto como ley universal. Pero no sucedería lo mismo al contrario: aunque llegáramos a la conclusión de que todos los fenómenos de la neurociencia son deterministas, esto no descartaría la posibilidad de encontrar indeterminaciones en otros ámbitos. De lo que se trata aquí, por tanto, es de averiguar cómo puede la neurociencia ayudar a responder al dilema de si el determinismo es o no cierto. Si consideramos la evidencia arrojada por los estudios de lesiones, de neuroimagen funcional y de neuroelectrofisiología, observaremos que la actividad de ciertas estructuras nerviosas (en especial el CPF) es una *condición necesaria* para que tenga lugar el proceso volitivo, aunque, como ya señalamos, no es obligatoriamente una *condición suficiente*. En efecto, y si bien no hay duda de que la función volitiva tiene un correlato anatómico, ¿por qué descartar la posibilidad de que el funcionamiento de las estructuras involucradas esconda procesos indeterministas? No obstante, tampoco hay evidencia empírica en esta línea. Por otro lado, si consideramos la postura de Libet acerca de la volición, independientemente de las críticas recibidas, observaremos que este autor plantea la posibilidad de que el mecanismo consciente de veto sea la manifestación de nuestra libertad de decisión. Sin embargo, no explica si este mecanismo tiene lugar de modo determinista o indeterminista (Libet 1999, p. 55), por mucho que piense que este último es tan válido como el primero o mejor (*ibid.*, pp. 56-7). En cuanto a la opinión de Wegner, podemos clasificarla dentro del determinismo duro porque el autor equipara la voluntad al libre

albedrío y piensa que es una ilusión que puede ser explicada de modo determinista, pero no olvidemos que ha recibido serias críticas. Finalmente, si aceptamos un influjo causal con sentido descendente del entorno sociocultural sobre nuestra voluntad y nuestro funcionamiento nervioso, podría parecer a simple vista que este fenómeno puede ser explicado mediante el determinismo social, pero no podemos descartar en absoluto que, por ejemplo, indeterminaciones de tipo cuántico intervengan en fenómenos ecológicos o climatológicos que influyen en la configuración de dicho entorno, o que, dado que la sociedad y la cultura son creaciones humanas, una conjunción de indeterminaciones en diversos agentes intervenga en la producción de dichas creaciones. En definitiva, y teniendo en cuenta todo lo anterior, la respuesta a la Cuestión Determinista, a la vista de los fenómenos descritos por la neurociencia, está en el aire.



## Capítulo 4. Causalidad mental, indeterminismo y control último: ¿es posible un libertarismo en armonía con la ciencia?

«Nunca has tenido una cámara en mi cerebro.»

*El show de Truman (Una vida en directo)* (Peter Weir, 1998).<sup>1</sup>

En este capítulo estudiaremos las posibilidades del libertarismo de dar con una teoría del libre albedrío que esté en total armonía con los conocimientos aportados por la ciencia contemporánea. Dicha teoría debería pasar por el cumplimiento de, al menos, tres condiciones fundamentales:

- a) *Causalidad mental*. Para ejercer nuestra libertad resulta especialmente importante la causalidad intencional: la causalidad mental de los estados intencionales.
- b) *Correlación inteligible entre la causalidad mental y un sustrato material*. Dicha correlación no puede conducirnos al epifenomenismo,<sup>2</sup> irreconciliable por definición con la causalidad mental, sino que debe traer consigo un sentido descendente de la causalidad.

---

<sup>1</sup> El guión de esta película es obra de Andrew Niccol. El fragmento transcrito corresponde a la versión doblada al castellano.

<sup>2</sup> El epifenomenismo (ya sea el dualista o el monista), como vimos en la sección 3.1., es la postura según la cual el cerebro influye causalmente sobre la mente, pero no viceversa.

- c) *Indeterminación ontológica que permita el control último.* Dicha indeterminación, por tanto, debe resistir al argumento de *Mind*. Además, permitir el control último la hará relevante para la causalidad mental.

A lo largo del capítulo analizaremos en profundidad cada una de estas tres condiciones. Esto nos proporcionará la base para presentar, ya en el capítulo siguiente, una propuesta propia.

#### **4.1. Primera condición: causalidad mental**

Previamente al estudio de la causalidad mental, se antoja necesario dejar establecido un concepto general de causalidad. Se trata de un problema de gran envergadura, que merece un tratamiento especializado, de manera que para poder trabajar con el caso concreto de la causalidad mental nos conformaremos con entender la causalidad como una relación en la cual un evento A (la causa) hace que la aparición de otro evento B (el efecto) sea más probable que cuando A no tiene lugar. Esta concepción, propuesta por Carlos Moya (2011, pp. 186-7), tiene la ventaja de ser lo suficientemente amplia como para ser “compatible con distintas visiones de la causalidad, deterministas o probabilistas, nomológicas, contrafácticas, y otras” (*ibid.*, p. 187), aunque es indispensable que B ciertamente *ocurra* para que A sea causa de B (Moya, comunicación personal, 2014). Una vez fijado un concepto general de causalidad, podemos, ahora sí, comenzar a tratar sobre la causalidad mental.

El propio Moya ha tratado con gran claridad expositiva el debate acerca de la causalidad mental y, además, ha participado muy activamente en él. Por ello acudiremos a este autor como autorizado guía principal en el tratamiento de este problema. Moya ha definido la causalidad mental como “el proceso por el cual ciertos eventos o estados, *en virtud de ser mentales, i.e. de tener propiedades*

*mentales*, dan lugar a cambios en las propiedades físicas del mundo, provocando eventos o estados físicos tales como cambios cerebrales, contracciones de los músculos y acciones corporales manifiestas” (Moya 2011, p. 185). Dicho de forma breve, se trata de la influencia causal de la mente sobre lo físico a través de las características específicamente mentales. Es importante aclarar que quedan fuera de esta definición aquellas situaciones en las que un evento físico recibe un influjo causal por parte de otro evento que goza de propiedades mentales que, sin embargo, no son relevantes para provocar dicho cambio, pues en estas situaciones la causalidad es de tipo físico (*id.*; Moya 2006a, p. 208). En relación a esta aclaración, Dretske (1988, p. 79) ha ofrecido el revelador ejemplo de una soprano cuyo canto con un tono muy agudo es capaz de romper el cristal: el cristal se rompe por las vibraciones sonoras que transmiten el canto (que son una propiedad física), y no por el significado de este (que es una propiedad mental).

Moya ofrece una clasificación, con vocación orientativa, de los estados y las propiedades mentales, estableciendo cuatro categorías (Moya 2006a, pp. 18-19):

#### Grupos básicos

- Estados intencionales: intenciones, deseos, creencias, etc.
- Estados fenomenológicos: experiencias sensoriales, placer, dolor, etc.

#### Grupos derivados de los dos anteriores

- Estados mixtos: emociones y sentimientos.
- Disposiciones puras: inteligencia, envidia, generosidad, etc.

Podemos concebir la causalidad mental como el poder causal del contenido mental de cualquiera de los tres primeros grupos sobre el ámbito de lo físico.

La causalidad mental es, no cabe duda, indispensable para gozar de libre albedrío (Moya 2012, p. 110). Desde luego, ¿podemos pensar que un agente es libre si no es capaz de tener un control mental sobre sus actos? Sin embargo, hay cinco importantes tesis que parecen entrar en conflicto con la causalidad mental, a saber: (a) el anomalismo en el marco de lo mental, (b) el externismo semántico e intencional, (c) el funcionamiento estrictamente sintáctico del cerebro, concebido como el órgano responsable del comportamiento intencional, (d) el cierre causal de la física, y (e) la concepción jerárquica de la realidad que parte del nivel físico como metafísicamente fundamental [Moya 2006a, p. 209, para (a), (b) y (d); Moya 2011, p. 188, para (c), (d) y (e)]. Las tesis (a), (d) y (e) son aplicables tanto al estudio de los estados intencionales como al de los fenomenológicos [si bien (a) es más relevante en relación a los primeros], y por lo tanto podrían suponer una amenaza para la causalidad mental en general. Las tesis (b) y (c), en cambio, sólo podrían entrar en conflicto con la causalidad intencional. Veamos las cinco tesis a continuación.

#### *4.1.1. Argumentos en contra de la causalidad mental*

La tesis (a), el anomalismo de lo mental, sostiene que no hay leyes estrictas de carácter psicofísico (para la relación mente-cuerpo) ni de carácter psicológico (para la relación entre propiedades mentales), de manera que no hay manera de reducir lo mental al lenguaje de la física (Moya 2006a, p. 209). El defensor más importante de esta tesis ha sido Donald Davidson (véase su importante colección de ensayos *Essays on Actions and Events* en Davidson 2001), quien sostiene que los enunciados de carácter mental buscan aproximarse lo máximo posible, con criterios de racionalidad, al sistema íntegro de creencias y motivos de un agente, características estas que no comparten los enunciados de la física. Cabe añadir que este autor está de acuerdo con la concepción nomológica de la

causalidad, es decir, la creencia en que las relaciones de causalidad obedecen a leyes generales; lo que ocurre, sostiene, es que las leyes que conectan causalmente lo mental con lo físico son de índole puramente física, de forma que el anomalismo no impide que las razones sean causas de las acciones.<sup>3</sup> Según este argumento, el anomalismo no impediría la causalidad mental.

Ahora bien, recordemos que la causalidad mental consiste en la influencia causal de lo mental sobre lo físico, pero mediante sus características específicamente mentales. Dado que en la explicación de Davidson la relación causal entre mente y cuerpo, entre razones y acción, tiene lugar gracias a las propiedades físicas de lo mental, dicha explicación no sólo no da cuenta de la causalidad mental, sino que constituye una amenaza para esta: conduce a una concepción epifenoménica de la mente en la que las propiedades mentales pueden servir, eso sí, para justificar racionalmente las acciones, pero no para ejercer una influencia causal sobre lo físico (Moya 1998b; 2004, pp. 65-7; 2006a, pp. 210-11).

Davidson ha intentado argumentar contra la posición según la cual, en su explicación del anomalismo, desemboca en el epifenomenismo (Davidson 1993), y por otra parte ha habido propuestas que apuestan por la causalidad nomológica pero niegan el anomalismo (*e.g.*, Audi 1993), surgiendo problemas en ambos casos (Moya 2006a, pp. 211-13). Podemos hallar una posibilidad alternativa apartándonos de la causalidad nomológica, manteniendo a la vez el anomalismo, de manera que las razones ejercerían su influencia causal sobre las acciones a través de una relación normativa, y no mediante leyes de carácter científico, aunque nos encontraríamos con la dificultad de tener que negar el cierre causal de la física (Moya 1998b; 2006a, pp. 213-14), tesis que hemos etiquetado como (*c*) y de la que hablaremos en breve. No obstante, y a pesar de dicha dificultad, abrazaremos esta posibilidad ofrecida por Moya en nuestra propuesta (véase capítulo 5).

---

<sup>3</sup> El vínculo entre razones y acción se tratará en el próximo apartado.

La tesis (b), el externismo semántico e intencional, es opuesta al internismo semántico e intencional, el cual consiste en la convicción de que la naturaleza tanto del significado de los signos como de los estados intencionales de un agente es dependiente de factores únicamente internos a él (materiales o no); dicho de otro modo, dependen constitutivamente de estos, haya o no una dependencia causal (cf. Moya 2006a, pp. 153-4). Un internista consideraría que “un cambio en las causas externas sólo genera un cambio en el contenido [de nuestros estados mentales] si es acompañado por un cambio en los factores internos”, mientras que “un cambio en los factores internos determinantes del contenido genera un cambio en éste aun cuando no vaya acompañado de un cambio en el entorno externo” (Moya 2006a, p. 154). Los partidarios del externismo sostienen, por contra, que ni el significado de los signos ni los estados intencionales están constitutivamente supeditados a factores internos en exclusiva, sino que también se encuentran en situación de dependencia constitutiva respecto a realidades externas al agente, que pueden ser de tipo físico o social (cf. Moya 2006a, pp. 153, 161 y 214).

Tanto el internismo como el externismo presentan dificultades que tienen que ver con la explicación que ofrecen para diferentes dimensiones del contenido mental (Moya 2006a, pp. 152-3). Estas dimensiones (*ibid.*, p. 152) son la semántica (cómo el contenido alude a la realidad en comunión con el lenguaje), la explicativo-causal (de qué manera puede conducirnos a ejecutar acciones) y la epistemológica (cómo el conocimiento directo de sí mismo por parte del agente le permite el control crítico de sus acciones y sus creencias, convirtiéndolo en responsable de ellas). Mientras que el externismo explica más satisfactoriamente la dimensión semántica, el internismo da cuenta mejor de las otras dos, de manera que no parece fácil ofrecer una explicación unificada de las tres dimensiones apostando exclusivamente por una u otra posición (*ibid.*, p. 153).

Por lo que se refiere a la dimensión epistemológica, es más fácilmente inteligible un adecuado acceso directo de un agente a sus contenidos

intencionales desde una posición internista, pues esta no acude a un acceso empírico al mundo exterior, susceptible de desembocar en errores (*ibid.*, p. 157). En cuanto a la dimensión semántica del contenido de los estados intencionales, las razones por las cuales el externismo aplica mejor que el internismo quedan perfectamente plasmadas por Moya de la siguiente manera (*ibid.*, p. 166):

[E]l externismo integra adecuadamente la dimensión semántica e intencional del contenido al reconocer el carácter relacional del significado de los signos, ya sean emitidos, escritos o meramente pensados. La capacidad de los signos de representar la realidad y de denotar entidades distintas de ellos mismos, así como la capacidad de nuestro pensamiento de apuntar intencionalmente al mundo y representárnoslo, no resulta, en principio, misteriosa para una concepción, como el externismo, que concibe desde el principio el significado y el contenido como propiedades relacionales, dependientes constitutivamente del entorno externo en el que un individuo se halla inmerso. En cambio, el internismo [...] tiene serias dificultades para dar cuenta de dichas capacidades, puesto que debe entender el significado y la intencionalidad como propiedades puramente intrínsecas de un individuo, independientes de sus relaciones con el mundo. En este marco, las propiedades semánticas del contenido, incluido su propio carácter intencional, se tornan misteriosas o, cuando menos, difícilmente explicables.

En relación a la dimensión explicativo-causal, resulta fácil de acomodar *a priori* a una postura internista, en la cual el comportamiento de un agente depende, explicativa y causalmente, sólo de propiedades del contenido mental definidas constitutivamente por factores internos, siendo estos de índole física si se opta por un enfoque materialista (*ibid.*, p. 155). El externismo, por el contrario, se encuentra con una dificultad: si el contenido semántico e intencional no viene dado constitutivamente por las características internas, ¿cómo puede influir causalmente en el comportamiento de forma inteligible? (*ibid.*, p. 167). Pues bien, es justamente en esta dificultad del externismo semántico e intencional para dar cuenta de la dimensión explicativo-causal del contenido mental donde reside la razón de su conflicto con la causalidad mental. En la propuesta que presentaremos en el próximo capítulo, no obstante, argumentaremos en contra de dicho conflicto.

La tesis (c), el funcionamiento estrictamente sintáctico del cerebro, concebido como el órgano responsable del comportamiento intencional, consiste en afirmar “que los procesos y estructuras cerebrales sólo pueden ser sensibles a propiedades físicas o, a lo sumo, sintácticas de signos y eventos, no a sus propiedades semánticas, como el contenido o el significado” (Moya 2011, p. 190). En consecuencia, y suponiendo que el comportamiento voluntario esté dirigido por el cerebro, el contenido semántico constituiría un epifenómeno y no influiría causalmente en el comportamiento, siendo imposible la causalidad intencional (*id.*). Nos hallamos, por tanto, ante una tesis que puede servir como argumento para apoyar, desde una perspectiva materialista, el internismo en relación a la dimensión explicativo-causal del contenido mental.<sup>4</sup>

La amenaza que la tesis del funcionamiento estrictamente sintáctico del cerebro supone para la causalidad intencional puede ilustrarse a partir del funcionamiento no semántico de un ordenador (Moya 2006a, pp. 215-16):

Un ordenador responde exclusivamente a propiedades físicas de los signos, por ejemplo a su forma, pero no a sus propiedades semánticas o intencionales. Un ordenador programado para informar de las existencias de cada producto en un almacén de droguería es absolutamente insensible a lo que significan los signos que introducimos en él y los signos que emite: somos nosotros, no el ordenador, los que interpretamos una determinada secuencia de signos como referida, digamos, al número de paquetes de detergente; de hecho, el ordenador haría exactamente lo mismo si lo utilizásemos para controlar las existencias de productos completamente distintos. No parece físicamente posible construir un ordenador sensible al significado. Lo que sí se puede hacer, y eso es lo que hacen los ingenieros y los programadores, es

---

<sup>4</sup> Conviene recalcar que lo que rechaza la tesis (c) no es el externismo propiamente dicho, sino la efectividad causal de las propiedades semánticas. Como ya se ha comentado, el externismo parece armonizar mejor con la dimensión semántica del contenido mental, y es en este punto donde encontramos un conflicto entre la tesis (c) y el externismo, pero dicha tesis rechazaría también la influencia causal de contenidos semánticos descritos bajo un prisma internista. La teoría del significado de Frege, por ejemplo, se enmarca en el internismo semántico y, dicho sea de paso, esquivando satisfactoriamente los problemas de la no referencia y de la correferencia, que constituyen un importante reto para las teorías externistas del significado, aunque presenta también dificultades propias (Moya 2006a, pp.169-74). Si, pongamos por caso, quisiéramos conciliar la semántica fregeana con la dimensión explicativo-causal del contenido mental, nos veríamos obligados a lidiar con la tesis (c), al igual que el externismo semántico.

conseguir que el ordenador *simule* ser sensible al significado, que actúe *como* si fuera sensible al significado de los signos, correlacionando las diferencias de significado con diferencias sintácticas o físicas en los signos, de modo que el ordenador responda a éstas como si respondiera a aquéllas. En la medida en que una diferencia de significado no tenga correlato físico alguno en el ordenador, esa diferencia será absolutamente ignorada y no tendrá ninguna consecuencia en su funcionamiento.

En contraste con los ordenadores, destaca Moya (p. 216), los humanos estamos seguros de poseer la capacidad real de responder al significado de los signos. Sin embargo, si el cerebro es un sistema físico que dirige nuestra conducta y sí, como queda patente en el caso de los ordenadores, parece que no es posible que un sistema físico responda al contenido semántico de los signos, ¿no estará el comportamiento provocado estrictamente por causas de naturaleza sintáctica y física? (*id.*). Si esto es así, el ser humano, pese a ser más complejo, actúa como un ordenador en tanto en cuanto no responde al significado (*id.*). A pesar de estas dificultades, no obstante, sugeriremos en nuestra propia propuesta una manera mediante la cual el cerebro podría responder al contenido semántico, negando por tanto que dicho órgano trabaje sólo sintácticamente.

Refirámonos ahora a la tesis (*d*), o sea, el cierre causal de la física. Esta tesis defiende que “todo cambio físico tiene una explicación física completa” (*ibid.*, p. 209). Papineau (2002, pp. 44-6) la denomina “completitud de la física” y la defiende de esta manera (*ibid.*, p. 45):

Una vez hemos establecido un significado definitivo para ‘físico’, como equivalente a ‘inanimado’, por así decir, ¿no es sólo entonces una cuestión de sentido común que todos los efectos físicos tendrán causas físicas? En particular, si tomamos los efectos físicos en este sentido que normalmente atribuimos a causas conscientes, ¿no es obvio entonces que estos efectos pueden en principio ser siempre explicados completamente en términos de historias [...] físicas, acarreado el movimiento de materia (en brazos), procesos moleculares (en músculos), la acción de neurotransmisores (en cerebros)... etcétera?

Aceptar el cierre causal de la física implica, por tanto, que el dominio físico garantiza una explicación necesaria y suficiente para todos los eventos físicos, mientras que rechazarlo supondría solamente que existe una explicación física necesaria para cada evento físico (Bishop 2011a, p. 95). Por consiguiente, la incompatibilidad entre el cierre causal de la física y la causalidad mental (incluyendo la intencional) se nos presenta así: si admitimos que la mente influye causalmente sobre la conducta gracias a sus propiedades específicamente mentales (como puedan ser los contenidos de creencias, deseos y propósitos), estamos afirmando que no hay una explicación física completa para eventos de carácter físico, es decir, que no se cumple el cierre causal de la física (Moya 2006a, p. 222; 2011, pp. 188-9). Incluso si la conducta no llega a ser ejecutada pero la probabilidad de que tenga lugar ha aumentado gracias al contenido mental, nos encontraríamos ante una injerencia de la mente sobre factores de índole física, incumpléndose la tesis de la completitud de la física (Moya 2011, p. 189). Imaginemos, por ejemplo, que un hombre, basándose en el deseo de satisfacer los gustos de su mujer, decide por la mañana que esa misma noche la invitará a cenar, pero que, desgraciadamente, justo antes de salir de casa en dirección al restaurante sufre un accidente doméstico que le obliga a cambiar de planes y acudir al hospital. Indudablemente, la probabilidad de que el hombre saliera a cenar con su mujer esa noche había crecido en el momento de la mañana en que así lo decidió, aunque finalmente no acudiera al citado restaurante. En una situación como esta podemos considerar que no se cumple el cierre causal de la física, aunque la acción planeada no haya llegado a ejecutarse. Llegamos, así, a la conclusión de que desmarcarse de la tesis (d) “es admitir que puede no haber en principio una teoría física completa de los fenómenos físicos, que la física teórica, en la medida en que aspira a ser una teoría completa, debe dejar de ser pura física e invocar poderes causales irreductiblemente no físicos” (Kim 1993, p. 209).

La aceptación del cierre causal de la física trae consigo la aceptación de la tesis (e), la concepción jerárquica de la realidad que parte del nivel físico

como metafísicamente fundamental (Moya 2006a, p. 223), la cual se entiende como sigue (*id.*; Moya 2011, pp. 189-90). La realidad se encuentra dividida, desde una perspectiva ontológica, en diversos niveles de complejidad creciente.<sup>5</sup> La secuenciación de los distintos niveles responde a un criterio de orden temporal y de causalidad. Esto significa que la realidad física fue la primera en aparecer cuando surgió el universo y que esta originó la realidad descrita por la química; seguidamente, la realidad química dio paso a la biológica, y esta condujo a la mental, la cual trajo consigo la social. Basándonos en este doble criterio, cronológico y causal, los niveles físicos constituyen las capas metafísicamente fundamentales del universo, convirtiendo a la física en la ciencia básica por excelencia, capaz de explicar en último término la eficacia causal de los niveles estudiados por el resto de ciencias, mientras que no es posible lo contrario. Por lo tanto, la tesis (*e*) implica que la relación causal entre niveles tiene lugar, necesariamente, en un sentido ascendente. Dado que la causalidad intencional, y la causalidad mental en general, consiste en el influjo causal de lo mental sobre el mundo físico gracias a las propiedades específicamente mentales, acarrea una relación causal descendente, por lo que resulta inequívoca su incompatibilidad con la concepción jerárquica de la realidad que parte del nivel físico como metafísicamente fundamental. En la propuesta que presentamos en el capítulo 5 se asume esta incompatibilidad y, de hecho, se argumenta en contra de la tesis (*e*) y a favor de la causalidad descendente que tiene lugar cuando opera la causalidad mental, lo que supone, asimismo, negar que se cumpla la tesis del cierre causal de la física.

---

<sup>5</sup> Una posible división, como ya se comentó en la sección 3.6., es la que propone Ellis (2009, p. 64), quien enumera ocho niveles en función de las disciplinas que los estudian: (1) física de partículas, (2) física atómica, (3) química, (4) bioquímica, (5) biología celular, (6) fisiología, (7) psicología, y (8) sociología, economía, política.

#### 4.1.2. Causalidad intencional: ¿existe un vínculo causal entre razones y acción?

El apartado anterior describía el conflicto que tiene lugar cuando confrontamos la causalidad mental con cinco importantes tesis. Todas ellas tienen un problema de compatibilidad con la causalidad intencional, que es el tipo de causalidad mental que tiene lugar por medio de las propiedades de estados intencionales (como intenciones, deseos, creencias...). Así, “[e]n la causalidad intencional, las propiedades mentales de los eventos o estados causantes son propiedades semánticas, especialmente significados y contenidos” (Moya 2011, p. 187). Pues bien, una pregunta que ha suscitado gran preocupación entre los filósofos es la de si encontramos un vínculo entre la acción intencional y las razones para llevarla a cabo. Dicho de otra forma: ¿pueden ser las razones causas de la acción intencional?

Debemos aclarar, antes que nada, que resulta problemático precisar el significado exacto del concepto “acción intencional”.<sup>6</sup> En cualquier caso, el debate filosófico en torno a la pregunta que acabamos de formular ha sido muy vivo desde que Elizabeth Anscombe (1957) publicara su influyente libro *Intention*. En esta obra, la autora caracteriza la acción intencional de la siguiente forma: “¿Qué distingue las acciones que son intencionales de aquellas que no lo son? La respuesta que sugeriré es que son las acciones para las cuales se aplica un cierto sentido de la pregunta ‘¿Por qué?’; el sentido es, por supuesto, aquel en el cual la respuesta, si es positiva, da una razón para actuar” (Anscombe 1963, p. 9). Anscombe propone, como vemos, que si es adecuado preguntar por las razones de una acción, entonces dicha acción es intencional. Esta concepción, bien asentada, es la que empleamos aquí.

Existen teorías tanto causales como no causales de la acción intencional. Las teorías no causales sostienen que las razones no ofrecen una explicación de las acciones en términos de causalidad sino únicamente en

---

<sup>6</sup> Véase Moya (2006a, pp.189-92) para una breve exposición sobre este asunto.

términos de justificación. Como explica Moya (2006a, pp. 194-5), el principal argumento esgrimido en favor de esta postura es el argumento de la conexión lógica, según el cual:

- (a) La causalidad ha de ser vista desde una perspectiva humeana: la causa y el efecto guardan un vínculo que no consiste en una dependencia conceptual, sino en una particularización de una regularidad, comprobable empíricamente, en la sucesión temporal de dos clases de eventos.<sup>7</sup>
- (b) Razones y acción guardan un vínculo de dependencia conceptual, y no achacable a la realización particular de una regularidad comprobable empíricamente.
- (c) Por lo tanto, las razones no causan la acción.

Las teorías causales de la acción, por su parte, defienden que las razones sí pueden ser causas de la acción. Entre estas teorías destaca la desarrollada por Donald Davidson, de cuyas obras al respecto sobresale especialmente su artículo *Actions, Reasons, and Causes* (Davidson 1963). En este trabajo, el autor defiende la idea de que las razones, en concreto las razones primarias, ejercen una clase de influencia causal. Para que esta influencia tenga lugar son necesarias dos condiciones. La primera de ellas dice así: “C1. R es una razón primaria por la cual un agente ejecutó la acción A bajo la descripción d sólo si R consiste en una actitud favorable del agente hacia acciones con una cierta propiedad y en una creencia del agente en que A, bajo la descripción d, posee esa propiedad” (*ibid.*, p. 687). Como vemos, C1 describe cómo una acción con una determinada descripción y unas determinadas propiedades queda justificada en virtud de razones primarias, que consisten en la conjunción de una actitud favorable y una creencia con respecto a una acción con dicha descripción y dichas propiedades. Sin embargo, no basta con la justificación de la acción en virtud de razones primarias para que estas sirvan como explicación

---

<sup>7</sup> Obsérvese que la causalidad humeana es una causalidad nomológica.

de tipo causal, ya que “una persona puede tener una razón para una acción, y ejecutar la acción, y sin embargo esta razón puede no ser la razón por la cual la ejecutó. Para la relación entre una razón y una acción que esta explica es primordial la idea de que el agente ejecutó la acción *porque* tenía la razón” (*ibid.*, p. 691). Así, Davidson añade una segunda condición a su propuesta: “C2. Una razón primaria para una acción es su causa” (*ibid.*, p. 693).

Según destaca Moya (2006a, pp. 202-3 y 205), hay dos grandes amenazas a la teoría causal de Davidson: el epifenomenismo y las cadenas causales desviadas. En el apartado precedente ya avanzábamos la amenaza del epifenomenismo. Recordemos que Davidson coincide con la concepción nomológica de la causalidad, según la cual las relaciones de causalidad obedecen a leyes generales, pero sostiene que las leyes que conectan causalmente lo mental con lo físico son leyes físicas. Por lo tanto, *el contenido mental* de las razones puede justificar racionalmente las acciones pero no puede causarlas, lo que constituye una concepción epifenoménica de los contenidos mentales (Moya 1998b; 2004, pp. 65-7; 2006a, pp. 205 y 210-11).

En cuanto a las cadenas causales desviadas, ofrecemos aquí, de entre los muchos ejemplos existentes, uno propuesto por Frankfurt (1978, p. 157): “[U]n hombre en una fiesta tiene pensado derramar lo que hay en su vaso porque quiere hacer señales a sus cómplices para comenzar un robo y cree, en virtud de sus acuerdos previos, que derramar lo que hay en su vaso logrará eso; pero todo esto conduce al hombre a estar muy ansioso, su ansiedad hace temblar su mano, y de esta manera su vaso se derrama”. El ejemplo cumple la condición C1 de Davidson: la acción intencional de derramar lo que contiene el vaso en las condiciones exactas en que lo había planeado el agente está justificada por la razón primaria consistente en la conjunción de una actitud favorable y una creencia con respecto a una acción de derramamiento ejecutada en esas mismas condiciones. También cumple la condición C2: la razón primaria de derramar el líquido en esas condiciones concretas acaba causando la acción de derramar. Pero, y he aquí el problema, la razón primaria causa un

acto de derramamiento con descripción y propiedades diferentes de las que conforman el acto de derramamiento planeado. En ese sentido concreto, podemos afirmar que el agente no derramó el contenido del vaso de forma intencional (cf. Moya 2006a, pp. 202-3, para una argumentación equivalente a esta, si bien con mayor profundidad y empleando ejemplos distintos). Similar es el razonamiento que emplea Frankfurt (1978, p. 157): “[E]l derramamiento [...] tiene entre sus causas un deseo y una creencia, que racionalizan el derramamiento por parte del hombre de lo que hay en su vaso, pero el derramamiento tal y como ocurre no es una acción”.

Moya sostiene que las amenazas que tanto el epifenomenismo como las cadenas causales desviadas suponen para la teoría causal de Davidson surgen si se conciben la justificación y la causalidad como independientes entre sí (concepción que está presente en las teorías causales y en las no causales), y, así, propone una posible solución consistente en la fusión de ambas, de manera que la justificación de una acción intencional sea al mismo tiempo su causa (Moya 1998a, 1998b; 2006a, pp. 203-5). En síntesis (2006a, p. 204):

El núcleo esencial de nuestra propuesta es el siguiente: si una razón no causa una acción, la acción en cuestión no está justificada por dicha razón. Así, la única condición que se requiere para que una acción sea intencional es que dicha acción esté justificada, bajo la descripción adecuada, por las razones del agente. El requisito de causalidad ya está incluido en la condición de justificación.

Se trata de un planteamiento que lleva consigo la aceptación de la causalidad mental, pues propone que ciertos contenidos mentales (los contenidos de las razones) influyen causalmente en ciertos eventos físicos (las acciones) (*ibid.*, p. 206). Como vimos en el apartado anterior, la causalidad mental se enfrenta a cinco importantes tesis, pero esto no será impedimento para que abracemos la solución ofrecida por Moya en la propuesta propia que aparece en el capítulo 5.

Se podría pensar que al analizar la problemática referente a la causalidad intencional dejamos de lado a las teorías no causales de la acción intencional, pero podemos argumentar que eso no es cierto (Moya 2011, pp.

187-8). Volvamos al concepto de causalidad, propuesto por Moya, y que ofrecíamos nada más comenzar esta sección: una relación en la cual un evento A (la causa) hace que la aparición de otro evento B (el efecto) sea más probable que cuando A no tiene lugar, siendo indispensable que B ocurra. Ahora planteemos la siguiente pregunta a un partidario de la no causalidad: ¿las razones de un agente para ejecutar una acción incrementan la probabilidad de que la acción se ejecute? Una respuesta negativa haría, posiblemente, que la justificación racional de la acción perdiera su sentido, porque ¿cómo justificaríamos una acción en virtud de ciertas razones si en ausencia de dichas razones la ejecución de la acción es igual de probable que en su presencia? Si la respuesta fuera positiva, lo cual parece más razonable, este partidario de la no causalidad de las razones estaría aceptando que las razones influyen de alguna manera en el ámbito de lo físico, siendo conveniente, por tanto, que tratara de aclarar cómo opera esa influencia. Así, “si tenemos razón en que aceptar el supuesto anterior acerca de la causalidad es suficiente para que surja el problema de la causalidad mental, entonces la causalidad mental [y en concreto, la causalidad intencional], con un nombre distinto, es también un problema para los no-causalistas” (*id.*).

#### *4.1.3. Papel de los estados fenomenológicos en la causalidad mental*

Finalizaremos la presente sección con algunas observaciones sobre el problema de la causalidad mental relativa a los estados fenomenológicos. Como vimos al principio de la sección, podíamos clasificar los estados mentales en dos grupos básicos (estados intencionales y estados fenomenológicos) y en dos grupos derivados (estados mixtos y disposiciones puras). Sobre la causalidad intencional, es decir, la causalidad mental de los estados intencionales, ya hemos hablado en los dos apartados precedentes. En cuanto a lo que podríamos llamar “causalidad fenomenológica”, o sea, la causalidad mental de

los estados fenomenológicos, cabe recordar que entra en conflicto, de la misma forma que la causalidad intencional, con tres de las cinco tesis que analizábamos en el apartado 4.1.1.: (a) el anomalismo en el marco de lo mental, (d) el cierre causal de la física, y (e) la concepción jerárquica de la realidad que parte del nivel físico como metafísicamente fundamental.

Si nos preguntamos si los estados fenomenológicos pueden ejercer por sí solos una influencia causal, gracias a sus características específicamente mentales, sobre el mundo físico, debemos regresar al conflicto generado por la interacción de las tesis mencionadas con la causalidad mental. Ahora bien, esto no implica que defender la causalidad intencional frente a las tesis (a), (d) y (e) obligue a defender la causalidad fenomenológica frente a esas mismas tesis. Uno puede ser partidario de un tipo de causalidad y no serlo del otro, o puede ser partidario de ambos, o no serlo de ninguno. No en vano,

[l]a diferencia entre los estados [intencionales y fenomenológicos] es muy notable y establece una dualidad en nuestro concepto de la mente. No parece haber un rasgo sustantivo común a una creencia determinada y a una sensación de dolor por el que clasifiquemos ambas como estados mentales. No parece haber, en suma, una esencia de lo mental, una característica esencial que defina todo aquello que es mental y lo distinga de aquello que no lo es (Moya 2006a, p. 19).

Así, por ejemplo, el problema de los *qualia* (los atributos de índole subjetiva de la experiencia individual, *e.g.*, el olor de un perfume, la blancura del blanco...) parece ir ligado exclusivamente a los estados fenomenológicos.<sup>8</sup> Pongamos por caso que se afirmara que los *qualia* son propiedades epifenoménicas de la mente; ello no obligaría a sostener que los estados intencionales tienen también una naturaleza epifenoménica.

---

<sup>8</sup> Esta afirmación no sería compartida por todos los filósofos. De hecho, hay una gran controversia sobre a qué estados mentales se asocian los *qualia*. También sobre cuál es su naturaleza e, incluso, sobre si existen realmente. El problema de los *qualia* merecería un extenso análisis, pues se trata de uno de los temas favoritos entre los filósofos de la mente, pero tal análisis se antoja excesivamente ambicioso para los objetivos de este trabajo.

Más relevante para el debate sobre el libre albedrío que la causalidad de los estados fenomenológicos en solitario parece la interacción entre dichos estados y los estados intencionales. Esta interacción puede tener lugar de distintas y complicadas formas (*ibid.*, p. 23). Una posibilidad es que la interacción dé lugar a emociones y sentimientos, es decir, a estados mixtos, en los que se halla una manifestación fenomenológica que está dirigida a un contenido intencional (*ibid.*, pp. 18-19). Sin duda, cuando tenemos un sentimiento o una emoción, los tenemos “hacia algo” o “sobre algo”:

[P]ensemos, por ejemplo, en suprimir del temor el objeto del mismo; lo que resta es una vaga intranquilidad que podría ser propia de emociones muy diversas. Este tipo de estados serían asimilables a los estados intencionales. En cambio, una depresión inespecífica o una tristeza sin objeto (quizá cabría denominar estas condiciones psíquicas «estados de ánimo») bien podrían agruparse junto con los estados fenomenológicos (*ibid.*, p. 19).

Pero, ¿pueden los estados mixtos operar causalmente sobre lo físico? Obviamente, se enfrentan a las dificultades que surgen de las cinco tesis sobradamente conocidas a estas alturas. No obstante, merece la pena tratar de profundizar algo más en el mecanismo de aparición de los estados mixtos a raíz de la convergencia entre estados intencionales y estados fenomenológicos. A este respecto, resulta enormemente interesante la hipótesis de los marcadores somáticos, desarrollada por Antonio Damasio, en especial en su libro *Descartes' Error* (Damasio 1994). Resumimos a continuación las ideas fundamentales de esta hipótesis, desgranada por su autor en el capítulo 8 del libro (*ibid.*, pp. 165-201).

Imaginemos que somos dueños de un negocio y tenemos la oportunidad de cerrar un buen trato con un potencial cliente que, desgraciadamente, está enemistado con nuestro mejor amigo. Tenemos que decidir si citarnos con él o no. En circunstancias normales, responderemos a esta situación concibiendo una serie de hipotéticos escenarios, representados mediante imágenes que se suceden unas a otras súbitamente. Una posibilidad de solucionar la situación consiste en proceder de manera puramente racional,

efectuando un estudio de los pros y contras que presenta cada uno de los escenarios. Desde luego, en muchas circunstancias de nuestra vida nos enfrentamos a más de dos opciones entre las cuales escoger, pero aun habiendo sólo dos, la situación resulta compleja (*ibid.*, p. 171):

Ganar un cliente podría suponer una recompensa inmediata y también una considerable cantidad de recompensa futura. Se desconoce cuánta recompensa y por lo tanto debes estimar su magnitud y tasa, a lo largo del tiempo, de modo que puedas enfrentarla a las pérdidas potenciales dentro de las cuales debes incluir ahora las consecuencias de perder una amistad. Puesto que esta última pérdida variará a lo largo del tiempo, ¿debes también calcular su tasa de “depreciación”? Te enfrentas, de hecho, con un cálculo complejo, situado en diversas épocas imaginarias, y cargas con la necesidad de comparar resultados de diferente naturaleza que de un modo u otro deben traducirse a una divisa común para que la comparación tenga algún sentido. Una parte importante de este cálculo dependerá de la generación continua de aún más escenarios imaginarios, contruidos sobre patrones visuales y auditivos, entre otros, y también de la generación continua de narraciones verbales que acompañan a esos escenarios, y que son esenciales para que continúe el proceso de inferencia lógica.

Ante este panorama nos enfrentamos a diferentes dificultades, entre las cuales se halla la limitación de nuestra memoria para conservar durante el proceso de razonamiento toda la información que necesitamos. Pero sabemos que, a pesar de todo, somos capaces de tomar decisiones en cortos intervalos de tiempo, para lo cual debe operar algo más que la razón.

Imaginemos, ahora, que previamente al proceso de razonamiento, notamos una breve sensación desagradable cuando imaginamos una consecuencia negativa de una posible decisión. Estamos ante un marcador somático, es decir, una sensación (visceral o no visceral) que procede de nuestro cuerpo y que se asocia a una imagen proyectada por nosotros. Respecto a su génesis, podemos decir que “*los marcadores somáticos son un caso especial de sentimientos generados a partir de emociones secundarias. Esas emociones y [esos] sentimientos se han conectado, mediante aprendizaje, a consecuencias futuras pronosticadas de ciertos escenarios?*” (*ibid.*, p. 174). No obstante, conforme avance nuestro desarrollo pueden aparecer símbolos de estados somáticos que

reemplazan a sensaciones reales en nuestro organismo. Por otra parte, los marcadores somáticos pueden ser negativos o positivos, según actúen como señales de alarma o como incentivos, respectivamente. También están presentes en aquellas situaciones en las que escogemos una opción que a corto plazo nos resulta negativa pero que a largo plazo nos traerá consecuencias positivas. Son gestionados a través del córtex prefrontal, y son capaces de actuar de manera consciente o inconsciente; en este último caso podrían ser el origen de nuestra intuición. Sea como fuere, los marcadores somáticos actúan como *asistentes del proceso de toma de decisiones* porque pueden conducirnos *ípsa facto* a descartar o a tener en cuenta una opción frente a otras. Eso sí, no tienen por qué bastar por sí solos para la toma de decisiones en cualquier situación, pues en muchos casos lo que hacen es permitir que un razonamiento posterior sea menos complejo de llevar a cabo. También puede suceder que los marcadores somáticos obstaculicen el razonamiento en algunas ocasiones.

La hipótesis de los marcadores somáticos, cuyos grandes rasgos acabamos de exponer, nos permite extraer algunas conclusiones interesantes. En primer lugar, parece proporcionar una solución satisfactoria para el problema del marco, el cual nos reta a explicar cómo nuestra inteligencia puede acceder de forma rápida y eficiente a la información significativa en la toma de una decisión, sin tener que analizar toda la información de la que dispone realmente.

En segundo lugar, da la impresión de que la hipótesis ofrece una base prometedora para explicar cómo pueden surgir estados mixtos de la mente a partir de componentes tanto intencionales como fenomenológicos. Damasio describe los marcadores somáticos como sentimientos derivados de emociones, de manera que, en el contexto en el que nos encontramos, podemos considerarlos quizá como equivalentes a los estados mixtos o, al menos, a muchos de ellos. De todos modos, esta base neurocientífica requeriría un adecuado complemento filosófico que explicara la relación entre estados físicos y estados mentales mixtos.

En tercer lugar, la hipótesis de los marcadores somáticos constituye una convincente integración de distintos protagonistas en el proceso volitivo. En el capítulo anterior (apartado 3.3.5.) destacábamos el triángulo sistémico que la cognición, la emoción y la memoria constituyen durante el proceso volitivo. Damasio propone una idea parecida, si bien en relación al razonamiento: “Hay por consiguiente tres papeles de apoyo en el proceso de razonamiento sobre un vasto panorama de escenarios generados a partir del conocimiento factual: *los estados somáticos automatizados*, con sus mecanismos de sesgo; *la memoria de trabajo*; y *la atención*” (*ibid.*, p. 198).

La cuarta conclusión que desprendemos es sostenida por el mismo Damasio; de hecho, es, posiblemente, la idea más importante que el autor defiende en *Descartes' Error*: el cerebro no actúa como un agente solitario y aislado del resto del cuerpo. Pensar lo contrario es caer en lo que Moya (2011, p. 198), muy elocuentemente, ha calificado como “una especie de dualismo materialista”:

Mientras que, de acuerdo con antiguas formas de dualismo, un ser humano es esencialmente un alma o cosa pensante que habita contingentemente un cuerpo, para este nuevo dualismo un ser humano es esencialmente un cerebro que habita contingentemente (el resto de) el cuerpo. En la filosofía de la mente actual, e incluso en la epistemología,<sup>9</sup> la importancia del cerebro, que no queremos negar, ha sido magnificada, con una correspondiente devaluación de otras partes del cuerpo humano, tales como la lengua o las manos, que tienden a aparecer como meros apéndices periféricos al servicio del cerebro, y en último término prescindibles. Todo tipo de actividad corporal, como los movimientos de las extremidades, se concibe siempre como un mero efecto de los procesos cerebrales, y nunca como una fuente o causa de lo que ocurre o existe en el cerebro.

Qué duda cabe de que la hipótesis de los marcadores somáticos supone un ataque contra este “dualismo materialista”. Su concepción unitaria (monista, si

---

<sup>9</sup> Nota de traducción: el término “epistemology”, que ha sido traducido aquí como “epistemología”, hace referencia a la teoría del conocimiento. No obstante, la influencia del inglés ha provocado que a menudo, y con una frecuencia creciente, se utilice “epistemología” como sinónimo de “teoría del conocimiento” en el ámbito hispanohablante.

se quiere) del organismo humano está refrendada, además, por importantes descubrimientos científicos. Por ejemplo, se sabe desde hace unas tres décadas que los sistemas nervioso, endocrino e inmune se encuentran intercomunicados a través de una red formada por neuropéptidos, que actúan como moléculas mensajeras, y sus correspondientes receptores (Pert *et al.* 1985). Lo más interesante de este descubrimiento es que los receptores están repartidos por dicha red de tal manera que podemos encontrarlos también en el cerebro, incluyendo regiones asociadas a las emociones. Este descubrimiento no hace sino reforzar la hipótesis de los marcadores somáticos, pues sugiere que el sistema nervioso es sensible a las señales emocionales procedentes de otros lugares del cuerpo.

Para cerrar el apartado, haremos una breve observación acerca de las disposiciones puras. De acuerdo con la clasificación de Moya que citábamos al principio de la sección, consisten en un grupo derivado de los dos grupos básicos (estados intencionales y fenomenológicos). Las disposiciones puras son capacidades y rasgos de carácter, cuya descripción requiere el uso de conceptos relativos a estados intencionales o fenomenológicos, y por eso conforman un grupo derivado (Moya 2006a, p. 19). Sin embargo, no tienen contenidos intencionales o fenomenológicos (*id.*).

#### **4.2. Segunda condición: correlación inteligible entre la causalidad mental y un sustrato material**

Cualquier teoría que sostenga que los seres humanos gozamos de libre albedrío (la libertarista entre ellas) necesita dar cuenta de la causalidad mental y, sobre todo, de la causalidad intencional, pues, al fin y al cabo, es el contenido semántico e intencional de algunos estados mentales el que goza del potencial para eventualmente causar las acciones que intuitivamente consideramos como

libres, por supuesto sin menoscabo de la influencia que los estados fenomenológicos puedan tener sobre dicho contenido. Sin embargo, hemos comprobado en la sección precedente que, desgraciadamente, no es sencillo ofrecer una explicación de la causalidad mental en armonía con la ciencia que sea completamente satisfactoria, una tarea que mantiene ocupada a una gran cantidad de filósofos desde hace mucho tiempo. Entre ellos, hay diversos autores que, a pesar de las dificultades, han intentado efectuar una aproximación materialista inteligible a la causalidad mental. Veamos a continuación algunos que resultan especialmente sugerentes.

#### *4.2.1. Causalidad mental y materialismo: en busca de una integración*

El primer intento que destacaremos es el de Fred Dretske (1988, 1993, 2009), quien ha elaborado una teoría informacional que aspira a la naturalización del contenido mental e intencional.<sup>10</sup> Para ello parte de una particular concepción del comportamiento. Si por este entendemos únicamente el movimiento corporal, entonces podemos dar cuenta de él atendiendo exclusivamente a *propiedades físicas* de estados internos. En este sentido, esos estados internos juegan el papel de “causas desencadenantes”, en palabras del autor. Pero el comportamiento no es movimiento, dice Dretske, sino aquellos procesos que conducen causalmente a él a partir de estados internos. La conexión causal se produce gracias a las propiedades físicas de estos, pero para saber por qué han ejercido el papel de causantes debemos atender a sus *propiedades semánticas*. En este sentido, dichos estados internos actúan como lo que el autor llama “causas estructurantes”. El significado que contienen los estados internos no es otra cosa que información acerca del entorno que se conecta con este por medio de

---

<sup>10</sup> La naturalización de la intencionalidad es la búsqueda de “una explicación de las propiedades intencionales y semánticas de la mente (en especial de la mente humana) sobre la base de propiedades que no son, a su vez, intencionales o semánticas” (Moya 1994, p. 228).

un vínculo causal de tipo nomológico establecido por aprendizaje. Sirviéndonos nuevamente de la terminología dretskeana, diremos que aquello que un estado interno *indica* respecto al entorno motiva su *reclutamiento* para la producción de un determinado movimiento (véase Dretske 1988, p. 101).

La teoría de Dretske se topa con algunos problemas.<sup>11</sup> El más destacado consiste en su incompatibilidad con la conocida como “causalidad aquí-y-ahora”. Este problema, que ha sido señalado por diversos autores (*e.g.*, Horgan 1991), puede describirse de la siguiente manera:

Nuestra intuición, dicen los críticos, no es tan sólo que nuestras creencias y deseos causan nuestros comportamientos, sino que lo hacen «aquí-y-ahora», es decir, que nuestros comportamientos tienen a creencias y deseos como antecedentes causales *locales*. Sin embargo, la teoría de Dretske ubica la eficacia causal de las propiedades indicadoras en un momento, seguramente bastante alejado, del pasado, cuando se dio aquel proceso de reclutamiento, por ejemplo, al finalizar el proceso de aprendizaje de un concepto (Vicente 2000, p. 11).

Aunque podría haber una escapatoria a este problema (véase Vicente 2000, pp. 11-13), no nos detendremos más con este asunto. Pese a que, como el lector comprobará más adelante, nuestra argumentación recibirá una marcada influencia dretskeana, consideramos que el problema de la causalidad “causalidad aquí-y-ahora” no resta interés al resultado de dicha argumentación. La razón es que este, que formalizaremos mediante dos tesis en el capítulo 5, alude, a diferencia de la propuesta de Dretske (quizá algo genérica en lo referente a sus aspectos biológicos) a un proceso neural concreto como participante en el proceso causal.

Pasemos a describir el intento de Robert Van Gulick (1993) de concebir la causalidad mental desde una perspectiva materialista. Las ciencias especiales, afirma este autor, tratan sobre objetos y eventos que están conformados por elementos físicos y que gozan de una influencia causal

---

<sup>11</sup> Al respecto, consúltese Moya (2006a, pp. 181-4 y 219; 2011, p. 196).

otorgada, en parte, por la conjunción de las propiedades físicas de estos con las leyes físicas, pero también por los *patrones organizativos* que los elementos establecen (*ibid.*, p. 250). Los predicados de las ciencias especiales singularizan estos patrones de organización (*id.*). Asimismo,

los patrones de orden superior pueden tener algún grado de independencia respecto de sus realizaciones físicas subyacentes y pueden ejercer lo que podría llamarse influencias causales descendentes sin requerir alguna forma objetable de emergentismo por la cual las propiedades de orden superior alterarían las leyes de la física subyacentes. Las propiedades de orden superior actúan mediante la *activación selectiva* de poderes físicos, no mediante su *alteración* (*ibid.*, p. 252).<sup>12</sup>

Según observa Carlos Moya (2006a, p. 226; 2011, p. 197), en la propuesta de Van Gulick es ciertamente inteligible que exista una activación selectiva en el ámbito de lo físico-químico procedente del nivel biológico, pero aún no lo es el proceso de activación selectiva por parte de lo intencional. Encontramos aún, por tanto, una distancia no trivial entre las explicaciones materialistas y las mentales e intencionales.

Finalmente, expondremos un reciente intento del propio Moya (2011, pp. 197-200) de armonizar el materialismo con la causalidad mental. Su aproximación otorga un marcado rol causal descendente a los contenidos intencionales.<sup>13</sup> Por otra parte, según afirma el propio autor (*ibid.*, p. 197), esta encuentra inspiración en las de Dretske y Van Gulick. Moya propone que pensemos en cómo los humanos nos insertamos en sistemas sociales de

---

<sup>12</sup> La propuesta de Van Gulick es compatible con el concepto de *emergencia contextual*, pues, a pesar de sus reticencias hacia “alguna forma objetable de emergentismo”, la emergencia contextual, como podremos comprobar en el próximo apartado, no implica que “las propiedades de orden superior alterarían las leyes de la física subyacentes”, y sí permite, en cambio, la activación selectiva de poderes causales en el nivel físico. Además, y tal y como opina Murphy (2011), la propuesta encaja muy bien con el papel del contexto y el funcionamiento de los sistemas complejos.

<sup>13</sup> Parte de las ideas sostenidas por Moya en su trabajo aparecen avanzadas en otro anterior (Moya 2006b, pp. 199-200), si bien son empleadas por el autor específicamente para defender un indeterminismo con sentido descendente que permite la responsabilidad moral.

carácter semántico, como puedan ser el lenguaje ordinario o la música. En situaciones de este tipo encontramos, sin duda, procesos de causalidad ascendente, ya que la producción de sonidos hablados o musicales, así como el progreso en las habilidades del habla o de la interpretación musical, no se explican sin la existencia de un sustrato causal de tipo neurofisiológico y un acrecentamiento en la complejidad de las conexiones neurales, respectivamente. Pero hay también un lugar para la causalidad descendente por medio de la selección y el refuerzo de determinadas conexiones neurales y neuromusculares:

[Q]ué conexiones neurales son seleccionadas entre la infinidad que son posibles depende del contenido semántico objetivo de los signos musicales o lingüísticos. Las conexiones que dan lugar a sonidos o hablas correctos son reforzadas durante el proceso de aprendizaje, mientras que aquellas que producen los resultados equivocados no lo son; y la diferencia está determinada por las reglas semánticas que gobiernan los sistemas. En este sentido, ciertas conexiones dentro del cerebro, así como [de este] con nervios y músculos, son seleccionadas y moldeadas mediante un proceso de causalidad descendente: de los contenidos y significados de los signos musicales y lingüísticos, de acuerdo a las reglas semánticas de la notación musical y el lenguaje ordinario, a las conexiones neurales y neurofisiológicas (*ibid.*, pp. 197-8).

Por tanto, continúa Moya, nos encontramos ante la influencia causal sobre el ámbito neural de factores (contenido y significado) que, conforme proclaman los externistas, trascienden el cerebro del agente. Además, la selección y el refuerzo que operan para configurar las conexiones neurales se producen a través de la selección de acciones corporales, de manera que es el cuerpo en su conjunto el protagonista del proceso, y no el cerebro en solitario. En definitiva, concluye el autor, da la impresión de que debe asumirse la causalidad mental e intencional para poder comprender y dar cuenta de situaciones como, pongamos, la sincronía que tiene lugar entre violinistas en el seno de una orquesta de calidad. En dicho caso

[d]ebe haber importantes patrones comunes en sus redes neurales que expliquen esa sincronía. Y estos patrones comunes no estarían ahí si procesos de causalidad descendente, partiendo de los significados objetivos de los signos musicales, no hubieran afectado a los cerebros, nervios y músculos de cada uno de esos sujetos mientras aprendían a leer la notación musical y a interpretarla con sus violines (*ibid.*, p. 199).

Ya se ha señalado anteriormente que Moya reconoce una inspiración a partir de la propuesta dretkeana. Desde luego, su planteamiento bien podría constituir una versión de dicha propuesta: determinadas conexiones neurales y neuromusculares son *reclutadas*, mediante su selección y su refuerzo, para ejercer funciones concretas debido a aquello que *indican* sobre los signos musicales y lingüísticos. Pensamos, eso sí, que la propuesta del autor da un paso adelante en la reducción de la brecha que separa las explicaciones materialistas de aquellas que se refieren a contenidos mentales. Moya *especifica* de un modo comprensible y coherente la capacidad de lo mental para activar selectivamente poderes causales subyacentes. Otra ventaja añadida de su planteamiento es que, al tener en cuenta la neuroplasticidad y la influencia causal del entorno sociocultural sobre la voluntad del individuo, se adapta a la perfección a la evidencia empírica sobre ambos fenómenos, de la cual ofrecimos una muestra en la sección 3.6.

Debido a las razones mencionadas, el intento de Moya nos servirá de influencia en el desarrollo de nuestra propia propuesta (véase capítulo 5). Sin embargo, no podemos fundar una posición libertarista y, al mismo tiempo, materialista sobre el libre albedrío únicamente en una justificación de la causalidad mental y la intencionalidad en términos de explicaciones materialistas, sino que, además, necesitamos encontrar una explicación inteligible de las posibilidades alternativas y, especialmente, del control último del agente, que es el control de sus decisiones y/o acciones compatible con el indeterminismo. Este extremo sobrepasa las ambiciones del trabajo donde Moya expone el intento que acabamos de describir. No obstante, el propio Moya (2006b, pp. 164-211) ha desarrollado una teoría cognitiva del control último (eso sí, sobre la responsabilidad moral) en la que, de nuevo, los sistemas

normativos juegan un papel decisivo, y en la que encontramos elocuentes ejemplos, siendo la notación musical uno de ellos. Hablaremos de esta teoría en el apartado 4.3.2. Cuando expongamos nuestra propuesta, empero, seguiremos una vía diferente a la de Moya en relación al control último, si bien, como ya hemos dicho, su intento de justificación de la causalidad intencional desde una perspectiva materialista nos será de inspiración para dicha tarea.

En este apartado hemos tratado con tres importantes intentos de integrar explicaciones mentales y materialistas en un único marco teórico. Estas aproximaciones (de Dretske, Van Gulick y Moya) nos serán de gran influencia a la hora de apostar por la causalidad descendente en general, y por la causalidad mental en particular. No obstante, para que nuestra argumentación resulte inteligible consideramos esencial compatibilizar dichas aproximaciones con una descripción adecuada de dos elementos fundamentales: *a)* un vínculo metafísico entre propiedades de niveles jerárquicos distintos que conlleve una relación causal descendente entre ellas, y *b)* un sustrato material a través del cual las propiedades de nivel superior operen sobre las propiedades de niveles inferiores. Los dos próximos apartados están dedicados, respectivamente, a ambas tareas.

#### *4.2.2. Reduccionismo, superveniencia y emergencia*

Recordemos la jerarquía de la realidad que ya hemos empleado como referencia (véase sección 3.6.), propuesta por Ellis (2009, p. 64) y compuesta por ocho niveles en función de las disciplinas que los estudian: (1) física de partículas, (2) física atómica, (3) química, (4) bioquímica, (5) biología celular, (6) fisiología, (7) psicología, y (8) sociología, economía, política. Atendiendo a ella, observaremos que la causalidad mental trae consigo un sentido descendente de la causalidad: consiste en el influjo causal de lo mental (correspondiente al nivel de la psicología) sobre el mundo físico por medio de propiedades

específicamente mentales. Por ello, la causalidad mental es incompatible, como ya vimos, con la concepción jerárquica de la realidad que parte del nivel físico como metafísicamente fundamental. Dicha tesis, obviamente, supone una concepción reduccionista del mundo.

A primera vista, apostar por el reduccionismo parece exactamente lo opuesto a hacerlo por la causalidad descendente. Sin embargo, deberíamos precisar qué entendemos por reduccionismo. Murphy (2011, pp. 202-3) distingue cuatro tipos: metodológico, epistemológico, ontológico y causal. El reduccionismo metodológico consiste en investigar una entidad en función de las partes que la componen. El reduccionismo epistemológico, por su parte, sostiene que las leyes y las teorías que dan cuenta de las capas inferiores de la realidad conducen a las que explican las capas superiores. En cuanto al reduccionismo ontológico, puede entenderse de dos formas. En una versión más débil, encontramos la opinión de que no hacen falta nuevos componentes de carácter ontológico (*e.g.*, una fuerza vital) conforme ascendemos por la jerarquía de la realidad. Esta versión, dice Murphy, es “inobjetable”, y estamos de acuerdo con ella; de hecho, en nuestra propia propuesta asumimos esta concepción reduccionista débil. Por lo que respecta a la versión fuerte del reduccionismo ontológico, “altamente objetable” para Murphy, y creemos que con razón, considera que sólo son ontológicamente reales los elementos que pertenecen al nivel inferior, mientras que los demás no son otra cosa que agregados perecederos de estos elementos. Finalmente, el reduccionismo causal mantiene que la manera en que actúan los componentes de una entidad determina lo que sucede en los niveles que se sitúan por encima, de forma que todas las relaciones causales tienen lugar en un sentido ascendente. Todo intento de abogar por un sentido descendente de la causalidad debe argumentar, por definición, en contra del reduccionismo causal. Este “se sigue de forma bastante directa del reduccionismo ontológico fuerte” (*ibid.*, p. 203), según el cual sólo son ontológicamente reales los elementos que pertenecen al nivel inferior, mientras que los demás no son otra cosa que agregados

perecederos de estos elementos. Por tanto, una defensa de la causalidad descendente debe comenzar por otorgar un estatus ontológico igualmente real a los elementos que conforman los demás niveles, siendo el nivel de la mente, lógicamente, el que más interés tiene para nosotros en este trabajo. Son dos las relaciones internivel que posibilitan una aproximación a este respecto: la superveniencia y la emergencia.

La superveniencia es una relación entre propiedades de dos niveles en la cual las propiedades del nivel más alto dependen de (están determinadas por) las propiedades del nivel más bajo, razón por la cual estas últimas constituyen la base subyacente de aquellas (es decir, son más básicas) (cf. Walter 2001, pp. 139-44; cf. Moya 2006a, pp. 100-101). Así, por ejemplo, las propiedades químicas supervienen a las propiedades físicas del átomo, las propiedades bioquímicas supervienen a las propiedades químicas, etc. De la misma manera, las propiedades mentales supervienen en última instancia a propiedades físicas. En la superveniencia, las propiedades del nivel inferior brindan condiciones *suficientes pero innecesarias* para alcanzar las del nivel superior, lo que conlleva una realización múltiple de estas últimas: cada una de ellas puede surgir a partir de diferentes bases subyacentes (Bishop y Atmanspacher 2006, p. 1757; cf. Moya 2006a, pp. 100-101).

La superveniencia no permite que tenga lugar una causalidad con sentido *descendente* entre dos niveles cualesquiera, como ocurre en la causalidad mental. Al menos, no vemos cómo podría hacerlo sin incurrir en una sobredeterminación causal. Dado que las propiedades de cada nivel vienen determinadas por las propiedades del nivel inmediatamente inferior, que son suficientes para llegar a ellas, la relación de superveniencia <sup>14</sup> marca un sentido

---

<sup>14</sup> Acerca de dicha relación, Kim (2006, p. 556) afirma: “[L]a superveniencia [...] no es una relación homogénea; no es una relación genuina, “natural”, sino más bien algo que surge a partir de relaciones naturales mantenidas a un nivel más profundo, como la causalidad y la reducción”. Sin duda, cabe el debate acerca de la existencia o inexistencia de la relación de superveniencia, así como de su naturaleza. Sin embargo, pensamos que dicho debate escapa a los objetivos de este trabajo, pues, sea como fuere, la superveniencia supone trazar un sentido

ascendente de causalidad. En cualquier caso, esto no conduce a negarle un estatus ontológico real a los niveles no físicos de la realidad, pues las propiedades de cada nivel no son idénticas a las propiedades del nivel precedente, sino que dependen de estas. Una relación de superveniencia, por tanto, no es una relación de identidad entre lo mental y lo físico (cf. Moya 2006a, p. 100). También conviene aclarar que el hecho de rechazar la aplicación de la superveniencia a la causalidad descendente (y, en consecuencia, a la mental) no significa negar su validez como vínculo internivel, ni siquiera entre lo físico y lo mental, en otros casos.

Descartada por nuestra parte la superveniencia como relación metafísica que permita una causalidad descendente, nos queda la opción que nos brinda la emergencia. Se trata de un concepto altamente problemático y no del todo consensuado, pues ha sido caracterizado de maneras diversas. No obstante, Bishop y Atmanspacher (2006, p. 1757) plantean dos clases de emergencia bien diferenciadas. He aquí una de ellas:

La descripción de propiedades en un nivel de descripción particular (incluyendo sus leyes) no ofrece condiciones *ni necesarias ni suficientes* para obtener la descripción de propiedades en un nivel superior. Esto representa una forma de *emergencia radical* desde el momento en que no hay condiciones relevantes que conecten en absoluto los dos niveles.

Debido a esta *desconexión*, la emergencia radical hace difícil, en primer lugar, dar cuenta satisfactoriamente de los estados mentales acudiendo a propiedades de elementos que pertenecen a niveles jerárquicos inferiores al mental, y, en segundo lugar, explicar de forma inteligible la comunicación de una influencia causal entre dos niveles (con sentido descendente, en nuestro caso). Desde luego, la emergencia radical se antoja poco compatible con nuestro conocimiento actual, según el cual debe existir algún tipo de dependencia entre propiedades de niveles sucesivos. Por todo ello, parece razonable descartar esta

---

ascendente, razón por la cual no puede dar cuenta de la causalidad mental, de carácter descendente.

primera variante de emergencia. Quizá resulte más adecuada una segunda alternativa, un concepto que ha venido desarrollándose durante los últimos años y que ha sido aplicado al estudio de la relación entre mente y cerebro (*e.g.*: Atmanspacher y beim Graben 2007, Atmanspacher 2012, Harbecke y Atmanspacher 2012). Se trata de la emergencia contextual:

La descripción de propiedades en un nivel de descripción particular (incluyendo sus leyes) ofrece condiciones *necesarias pero no suficientes* para obtener la descripción de propiedades en un nivel superior. Esta versión, que proponemos llamar *emergencia contextual*, indica que se requieren condiciones contextuales contingentes, además de la descripción de nivel inferior, para la obtención rigurosa de propiedades de nivel superior (Bishop y Atmanspacher 2006, p. 1757).<sup>15</sup>

El hecho de necesitar las propiedades inferiores para dar cuenta de las superiores marca una *conexión indeleble* entre niveles de descripción, lo que, en principio, da ventaja a la emergencia contextual, frente a la radical, a la hora de intentar explicar los estados mentales a partir de propiedades jerárquicamente inferiores a las mentales. Ahora bien, dado que el nivel inferior, aunque necesario, no resulta suficiente para dar con el nivel superior, para lo cual han de tenerse en cuenta las circunstancias contextuales, cabe afirmar que el nivel inferior no determina al superior. Por ello, es susceptible de recibir eventuales influencias causales procedentes de este último en virtud del contexto, de manera que no existen impedimentos para que la causalidad descendente pueda tener cabida en el seno de un vínculo de emergencia contextual.

El estudio de la emergencia, y/o de la importancia del ámbito relacional/contextual, en conjunción con la causalidad descendente ha sido una línea de investigación habitual en los últimos tiempos (véanse, por ejemplo, Bishop 2008, Ellis 2009, Juarrero 2009, Newsome 2009, Murphy 2011), y en

---

<sup>15</sup> Según esto, la descripción de un estado físico particular sería imprescindible para describir el estado mental relacionado. Por este motivo, la realización múltiple, en la cual “un determinado tipo psicológico (como el dolor) puede ser realizado por muchos tipos físicos distintos” (Bickle 2013), no sería posible con la emergencia contextual.

intonía con ella consideramos que la emergencia contextual constituye un excelente candidato como vínculo metafísico entre propiedades de niveles jerárquicos distintos que permita una relación causal descendente entre ellos. Pero describir dicho vínculo era sólo la primera tarea que nos habíamos impuesto con el fin de reforzar la inteligibilidad de nuestra argumentación. La segunda tarea, recordemos, consistía en describir un sustrato material a través del cual las propiedades de nivel superior operen sobre las propiedades de niveles inferiores. La búsqueda del cumplimiento de dicha tarea nos conduce a los sistemas complejos.

#### *4.2.3. Sistemas complejos y causalidad descendente*

Los sistemas complejos (o sistemas dinámicos complejos) vienen siendo ampliamente estudiados desde hace ya algún tiempo, pero, a pesar de ello, su definición exacta es notablemente diversa a la vista de la bibliografía al respecto (cf. Bishop 2008, p. 231; cf. Bishop 2011b, p. 111). No obstante, y siguiendo a Bishop (2008, pp. 232-3; 2011b, pp. 111-14), expondremos varias características fenomenológicas que se han empleado para describirlos:

(1) Necesitan varios componentes para manifestar comportamientos de tipo complejo (como puedan ser el aleatorio o el caótico). En algunos casos puede bastar con tres componentes.

(2) Son capaces de romper una simetría (*e.g.*, homogeneidad en la disposición espacial), de tal forma que esta no existirá después de que se atraviese un punto crítico.

(3) Están formados por una jerarquía reconocible de niveles estructurales. Es frecuente que un cambio de nivel conlleve un cambio de descripción. Asimismo, los componentes de niveles inferiores condicionan necesaria pero insuficientemente la presencia y la forma de actuar de algunas estructuras superiores. Existe también la posibilidad de que los niveles superiores influyan

en los componentes inferiores de modo tal que estos no condicionen necesaria y suficientemente su actuación propia. En esencia, dicha influencia tiene lugar mediante *constricción*, ya sea ambiental, de tipo interno, o ambas a la vez. Esta complicada estructura jerárquica conlleva un análisis matemático no lineal de su comportamiento.

(4) Con frecuencia están ligados a procesos de carácter irreversible, como por ejemplo la difusión molecular.

(5) Sus componentes no están agregados, sino interrelacionados.

(6) Es frecuente que la evolución temporal de sus componentes esté supeditada a la estructura que las contiene y a la historia y el ambiente de la totalidad del sistema.

(7) Poseen integridad funcional. No la encontraremos cuando falte una estructura o un componente interno o cuando desaparezca la interrelación estructuras-componentes.

(8) Sus componentes establecen diversos vínculos de tipo funcional/estructural (bucles de retroalimentación, por ejemplo) que ayudan a la conservación de la integridad sistémica.

(9) Su comportamiento es intrincado, a saber, no está ni en el orden ni en el desorden absolutos.

(10) Son estables: cuando sufren pequeñas perturbaciones conservan su unidad organizativa y relacional, y esta se adapta a modificaciones ambientales de carácter moderado.

(11) Están sujetos a relatividad observacional: el modo en que se observan y describen determina su complejidad.<sup>16</sup>

Tal y como el propio Bishop (2011b, pp. 128-9) observa, la emergencia contextual podría dar cuenta satisfactoriamente del funcionamiento de los sistemas complejos. Según dicho concepto, recordemos, las propiedades de un

---

<sup>16</sup> Bishop cita un par de trabajos (Grassberger 1989, Crutchfield 1994) de los que, efectivamente, podemos desprender el fenómeno mencionado. Por Nótese, por cierto, la marcada condición idealista de esta característica.

nivel brindan condiciones necesarias pero no suficientes para las propiedades de una escala superior. “Se requieren condiciones contingentes que especifiquen el contexto [en] la transición desde el nivel de propiedades y comportamientos inferior al superior para proporcionar tales condiciones suficientes”, señala el autor (*ibid.*, p. 128). Estas características de la emergencia contextual se adecuan, desde luego, a la configuración jerárquica de los sistemas complejos, en los que, como acabamos de señalar, los componentes inferiores son necesarios pero insuficientes para la presencia y la manera de actuar de algunas estructuras superiores, siendo además posible que los niveles superiores impongan *constricciones internas y/o ambientales* a los inferiores de tal forma que estos no condicionen necesaria y suficientemente su actuación propia.

Ya señalamos anteriormente que la causalidad descendente tiene cabida cuando existe un vínculo internivel de emergencia contextual. Si aceptamos que este vínculo se presenta en los sistemas complejos, entonces cabe decir que la causalidad descendente puede, en principio, operar en el seno de estos. Partiendo de este enfoque, nos disponemos ahora a efectuar una defensa de la causalidad mental en virtud de una integración entre la emergencia contextual y los sistemas complejos, por un lado, y las propuestas de Dretske, Van Gulick y Moya, por otro. Dicha integración será formalizada a través de tres tesis:  $\alpha$  (causalidad),  $\beta$  (causalidad descendente) y  $\gamma$  (causalidad mental).

Para empezar, resulta necesario precisar qué es la causalidad. Se trata de un problema de enorme calado, y merece sin duda un tratamiento especializado, de manera que, para poder trabajar aquí, nos conformaremos con entender la causalidad como una relación en la cual un evento A (la causa) hace que la aparición de otro evento B (el efecto) sea más probable que cuando A no tiene lugar, si bien es indispensable que B ciertamente *ocurra* para que A sea causa de B. Esta concepción, que ya presentábamos en la sección 4.1. y que debíamos a Moya, nos sirve para establecer nuestra tesis preliminar:

Tesis  $\alpha$  (CAUSALIDAD).

Un evento A (la causa) hace que la aparición de otro evento B (el efecto) sea más probable que cuando A no tiene lugar, siendo indispensable que B finalmente se produzca.<sup>17</sup>

Introduciendo en  $\alpha$  la ya mencionada propiedad de los sistemas complejos según la cual, en virtud de un vínculo de emergencia contextual, los niveles jerárquicos superiores pueden influir sobre los inferiores por medio de constricciones internas y/o ambientales, llegamos a la tesis que viene a continuación:

Tesis  $\beta$  (CAUSALIDAD DESCENDENTE).

La imposición de determinadas constricciones por parte del contexto interno y/o ambiental en un nivel jerárquico de un sistema complejo (la causa) hace que la activación de ciertos poderes causales en un nivel inferior (el efecto) sea más probable que en ausencia de esas constricciones, siendo indispensable que dicha activación finalmente se produzca.

Es menester destacar la influencia de las propuestas desarrolladas por Dretske y Van Gulick en el planteamiento de esta nuestra segunda tesis. En las dos hay

---

<sup>17</sup> Moya (comunicación personal, 2015) aclara que no estamos ante una caracterización exacta de la causalidad, sino ante una idea basada en una condición mínima necesaria para la existencia de una relación causal. No obstante, se han descrito ejemplos en los que la causa parece mermar la probabilidad de que el efecto se produzca, si bien este, de hecho, se produce. El siguiente es uno de ellos: “Imaginemos una pelota de golf rodando hacia el hoyo, de tal forma que su probabilidad de caer dentro es bastante alta. Por allí aparece una ardilla, que con una patada aleja la pelota del hoyo, reduciendo así su posibilidad de entrar. Entonces, a través de una serie de improbables rebotes, la pelota cae en el hoyo” (Sober 1984, p. 406). En este caso, la patada de la ardilla (la causa) parece mermar la probabilidad de que la entrada de la pelota (el efecto) se produzca, si bien esta, de hecho, entra en el hoyo. Permítasenos, pues, emplear nuestra descripción de la tesis  $\alpha$  como una mera aproximación a la relación de causalidad, siendo conscientes de que hay otras posibles.

una selección de poderes causales a escala física en la que el contexto juega un papel decisivo. Ahora bien, la naturaleza de dicho contexto difiere en función de la propuesta. Según Van Gulick, los *patrones* conformados a un nivel superior condicionan la actuación de los niveles inferiores. Estos patrones, recordemos, quedan establecidos mediante la organización de los elementos físicos, por lo que el contexto es de carácter *interno*.<sup>18</sup> Dado que el patrón (la forma) establecido internamente en un determinado nivel de un sistema constriñe el comportamiento de sus componentes, podemos decir que las constricciones mencionadas en  $\beta$ , aun no desempeñando el papel de causas eficientes aristotélicas, sí que actúan, en cambio, como *causas formales* (cf. Juarrero 1999, p. 146; cf. Juarrero 2009, p. 89; cf. Ellis 2009, pp. 77-8; cf. Bishop 2011b, p. 127; cf. Murphy 2011, pp. 214-15).<sup>19</sup>

Según Dretske, las *propiedades semánticas* de los estados internos condicionan la actuación de sus propiedades físicas. El significado, como dijimos, consiste en información respecto al entorno conectada con este por medio de un vínculo causal nomológico establecido por aprendizaje. Un estado interno es reclutado para la producción de un determinado movimiento por causa de aquello que indica sobre el entorno. En la propuesta de Dretske el contexto es de carácter *normativo*. Dado que dicho contexto, mediante su conexión con los niveles mental y el biológico, constriñe la actuación en el nivel físico, podemos decir que las constricciones mencionadas en  $\beta$ , aun no

---

<sup>18</sup> Van Gulick concibe el patrón como un contexto: “Un determinado constituyente físico podría tener muchos poderes causales, pero sólo algún subconjunto de ellos estará activo en una situación dada. *El contexto mayor (i.e., el patrón)* [las comillas son nuestras] del cual es parte podría afectar a cuáles de sus poderes causales quedan activados” (Van Gulick 1993, p. 251).

<sup>19</sup> Podríamos dudar acerca de si estamos en este caso ante un verdadero tipo de causalidad. Por ejemplo, Murphy (2011, p. 215) cree que el término “constricción todo-parte” es más adecuado que “causalidad descendente”. Permítasenos no tratar de profundizar en este asunto en tanto en cuanto no será una causalidad formal, sino otra de orientación dretskeana, la que nos marcará el camino hacia nuestra tesis de causalidad mental (tesis  $\gamma$ ). De cualquier modo, el lector puede hallar una aproximación a la física de la causalidad descendente en Davies (2006).

desempeñando el papel de causas eficientes (o desencadenantes), sí que pueden actuar, en cambio, a modo de *causas estructurantes dretskeanas*.<sup>20</sup>

En resumen, en nuestra tesis  $\beta$ , la noción general de causalidad contenida en  $\alpha$  ha quedado restringida acudiendo al papel del contexto en los sistemas complejos, ya sea este interno, que permitiría una causalidad descendente de tipo formal, o normativo, en cuyo caso estaríamos ante una causalidad descendente de tipo estructurante. Ahora bien, no basta con una tesis como  $\beta$  para obtener una explicación completa de la causalidad mental. Veamos por qué.

Existen análisis de sistemas complejos pertenecientes a diferentes niveles jerárquicos. Un caso físico-químico, ya clásico, es el de la convección de Rayleigh-Bénard (véase, por ejemplo, Bishop 2008). A escala biológica, Newsome (2009, p. 56) ha descrito lo que él denomina “comportamiento emergente” de los seres unicelulares. La existencia de dichos seres permite fenómenos como la búsqueda de recursos, la movilidad de la célula, la competencia ecológica, etc. Según Newsome, el mero conocimiento de las macromoléculas de estos organismos no faculta para la predicción de dichos fenómenos, y los principios que los rigen se desmarcan de los que rigen a aquellas. Además,

[c]ada uno de estos fenómenos debe ser identificado y descrito en y por sí mismo, y sus reglas lógicas internas [deben ser] comprendidas, antes de que puedan realizarse vínculos rigurosos a mecanismos de nivel inferior. Las interacciones competitivas entre especies, por ejemplo, son comprendidas mediante observación en el nivel conductual, no mediante inferencia a partir del nivel molecular. El comportamiento del organismo unicelular, a su vez, ejerce control descendente sobre sus moléculas constituyentes. El movimiento de un orgánulo dentro de la célula depende en cierto sentido de la presión ejercida desde el citoplasma mientras el organismo se mueve. Pero en otro sentido, igualmente válido, el movimiento del orgánulo depende del objetivo conductual inmediato del organismo. Hasta donde sabemos, nada acerca de la vida de los organismos unicelulares viola las leyes de la física o las

---

<sup>20</sup> La afirmación carecerá de sentido en los casos en los que no exista conexión semántica entre el sistema complejo y su entorno (por ejemplo, cuando el sistema sea no vivo).

leyes químicas que gobiernan el comportamiento de las macromoléculas. La célula no puede comportarse de cualquier manera que no esté permitida por los niveles inferiores de organización de sus partes constituyentes; el comportamiento de la célula está, así, *constrañido* pero no *determinado* por los niveles inferiores.

A nuestro modo de ver, el ejemplo ilustra la existencia de emergencia contextual y de causalidad descendente en el mundo vivo. En cuanto al ser humano, son varias las propiedades de los sistemas complejos que en él pueden observarse: comportamiento caótico (evolución impredecible tras una pequeña perturbación inicial, como una infección vírica), niveles jerárquicos (células, tejidos, órganos...), procesos irreversibles (*e.g.*, la difusión molecular en la membrana plasmática), vínculos funcionales que facilitan la integridad sistémica (es el caso de la retroinhibición endocrina), estabilidad (por ejemplo, la resistencia organizativa y relacional frente a lesiones o enfermedades de poca consideración), etc. Ahora bien, resulta obvio que la diferencia entre un ser unicelular y uno humano es enorme cuantitativamente (el grado de complejidad del organismo humano es infinitamente mayor) y, sobre todo, cualitativamente, puesto que el ser humano goza de contenidos mentales e intencionales. Eliminar o, al menos, reducir dicha diferencia cualitativa se antoja absolutamente imprescindible para que los sistemas complejos puedan ser considerados como el sustrato material a través del cual las propiedades mentales operan sobre las propiedades de niveles inferiores. Nuestra tesis  $\beta$  se halla ante una distancia no trivial entre las explicaciones materialistas y las mentales e intencionales, y esto es debido a que no discrimina la causalidad mental de otras variedades de causalidad descendente (cf. Moya 2006a, p. 226; cf. Moya 2011, p. 197). En nuestra opinión, la distancia mencionada disminuye en la aproximación efectuada por Moya, que ya presentamos en el apartado 4.2.1. Recordemos que, según este autor, en el progreso de las habilidades del habla o de la interpretación musical hay un lugar para la causalidad descendente por medio de la selección y el refuerzo de determinadas conexiones neurales y neuromusculares en virtud del contenido y el significado de los signos

lingüísticos y musicales, respectivamente. A nuestro juicio, su planteamiento es compatible con el concepto de emergencia contextual si pensamos en el contenido y el significado como constricciones contextuales de carácter normativo que completan la descripción necesaria pero insuficiente que, para explicar el desempeño de habilidades habladas o musicales, se extrae a nivel neural. Por ello, aplicaremos la aproximación de Moya a nuestra tesis  $\beta$  (causalidad descendente) para restringirla al caso específico de la causalidad mental, supeditando dicha aplicación a la concepción del sistema nervioso y de la mente como dos realidades jerárquicamente interrelacionadas (y propias de los niveles fisiológico y psicológico, respectivamente) en el seno de un mismo sistema complejo, un ser humano:

#### Tesis $\gamma$ (CAUSALIDAD MENTAL).

La imposición de determinadas constricciones de carácter semántico por parte del contexto normativo en la mente de un ser humano (la causa) hace que el proceso de selección y refuerzo de las rutas neurales y neuromusculares conducentes a la ejecución de ciertas acciones (el efecto) sea más probable que en ausencia de esas constricciones, siendo indispensable que dicho proceso finalmente se produzca.<sup>21</sup>

Cabría pensar en esta tercera tesis como una versión de la propuesta de Moya y, con ella, también de la de Dretske: el proceso de selección y refuerzo de rutas neurales y neuromusculares podría ser visto como un mecanismo de *reclutamiento* dretskeano.  $\gamma$  guarda también similitud con la propuesta de Murphy y Brown, para quienes los eventos mentales consisten en eventos cerebrales sumidos en el contexto de sistemas complejos que ellos denominan “bucles de acción-retroalimentación-evaluación-acción”, que actúan en el ambiente y que

---

<sup>21</sup> No pensamos en esta tesis como una opción excluyente de causalidad mental. Entendemos que son posibles también procesos de causalidad descendente en los que intervengan eventos y estados mentales de naturaleza no semántica, como los fenomenológicos.

gozan de poder causal sobre el cerebro por medio de constricciones (Murphy y Brown 2007; Murphy 2011, p. 211).<sup>22</sup> De cualquier forma, tesis como  $\gamma$  u otras semejantes adquirirían una potencia mayor si se lograra hallar una explicación neurobiológica concreta para el proceso de selección y refuerzo de rutas neurales y neuromusculares (“el efecto”) mencionado en  $\gamma$ . Sin duda, los fenómenos de neuroplasticidad, sobradamente documentados y acerca de los cuales ya hablamos en la sección 3.6., marcan una vía idónea para investigar en pos de dicha explicación. Nosotros sugeriremos una posibilidad basada en una forma particularmente intensa de neuroplasticidad: la poda sináptica. Pero eso será ya en el próximo capítulo. Ahora es momento de pasar al análisis de la tercera condición que estimábamos fundamental para hallar una respuesta libertarista y materialista al problema del libre albedrío.

#### **4.3. Tercera condición: indeterminación ontológica que permita el control último**

Cualquier teoría libertarista que se precie debe dar cuenta de cómo el indeterminismo permitiría al agente gozar de posibilidades alternativas y de control último sobre sus decisiones y/o acciones de un modo inteligible (véase sección 1.3.). Sin embargo, el argumento de *Mind*, según el cual el indeterminismo excluiría la libertad porque conlleva aleatoriedad y no existe control sobre algo que es aleatorio, supone un auténtico reto para el libertarismo. Como resultado, gran parte de los esfuerzos libertaristas se encaminan a intentar rebatir dicho argumento. Además, una indeterminación que permita el control último será relevante para la causalidad mental, en el sentido de que influirá en el poder causal de lo mental sobre lo físico. Por otra

---

<sup>22</sup> Véanse asimismo: Juarrero 2009, pp. 90 y 94-5; Newsome 2009, pp. 57-61.

parte, la indeterminación debe ser de tipo ontológico, de forma que refleje un aspecto esencial del universo, y no meramente epistemológico, lo que conlleva únicamente impredecibilidad (véase apartado 2.1.7.).

En esta sección atenderemos a dos propuestas, ofrecidas por Robert Kane y Carlos Moya, que constituyen, a nuestro parecer, muestras notablemente representativas de dos vías bien distintas de concebir el libre albedrío libertarista desde una perspectiva monista y materialista. Dada la temática de la presente sección, no buscaremos una caracterización exhaustiva, sino que nos centraremos en el modo en que cada una afronta el origen del indeterminismo y la relación de este con el control último.

#### 4.3.1. *Indeterminismo endógeno: la propuesta de Kane*

La contribución de Robert Kane al debate sobre el libre albedrío tiene un valor incalculable, tanto por su gran acervo en forma de publicaciones (*e.g.*: Kane 1985, 1989, 1996, 1999, 2000, 2002a, 2002b, 2005, 2006, 2007a, 2007b, 2008, 2009) como por su insistencia en la búsqueda de una armonía entre libertarismo y conocimiento científico. De un trabajo reciente (Kane 2011), extraeremos a continuación algunas ideas esenciales para comprender cómo concibe este autor el origen del indeterminismo y su relación con el control último.

El libre albedrío, sostiene Kane, es un tipo de libertad que no es compatible con el determinismo y que consiste en “el poder para ser el creador último y sustentador de los propios fines o propósitos” (*ibid.*, pp. 382-3).<sup>23</sup> Los propósitos, que son los contenidos de las intenciones, son creados por medio de decisiones o elecciones y sostenidos por medio de esfuerzos de voluntad. Ser creadores últimos de los propósitos es posible en virtud de poseer

---

<sup>23</sup> Esta definición aparece, de una forma casi idéntica, en Kane (1996, p. 15): “[E]l poder para ser *m* [cursiva añadida] creador último y sustentador de los propios fines o propósitos”.

responsabilidad última, o UR (“ultimate responsibility”), es decir, de “ser responsable por cualquier cosa que sea una razón suficiente (condición, causa o motivo) para que ocurra la acción” (*ibid.*, p. 383). Para tener UR no hace falta gozar de posibilidades alternativas, o AP (“alternative possibilities”), siempre que actuamos en virtud de nuestro libre albedrío, sino en relación con ciertas acciones o elecciones pasadas que conforman el carácter que tenemos en la actualidad. Se trata de las acciones autoformadoras, o SFAs (“self-forming actions”), las cuales están *indeterminadas* porque no hay motivos o causas suficientes para que se den. Las AP han de cumplir con las condiciones de voluntariedad, intencionalidad y racionalidad (denominadas por Kane “condiciones de pluralidad”) para que resulten interesantes a la hora de preguntarnos por la libertad de voluntad del agente volitivo. Las SFAs deben cumplir con dichas condiciones, lo cual conlleva cumplir con la condición de AP. Pero, además, las SFAs han de ser “will-setting”, lo que significa que la voluntad queda fijada hacia una dirección concreta, de modo intencional y voluntario, en el momento de ejecutarlas, y no con anterioridad a ellas.

Las SFAs tienen lugar en situaciones de conflicto entre distintas posibilidades motivacionales. Veamos el ejemplo de una mujer de negocios (*ibid.*, p. 387):

Va de camino a una reunión importante para su carrera cuando observa un asalto teniendo lugar en un callejón. Sucede una lucha interna entre su conciencia moral, para parar y pedir ayuda, y sus ambiciones profesionales que le dicen que no puede perderse esta reunión. Tiene que hacer un esfuerzo de voluntad para vencer la tentación de seguir hacia su reunión. Si vence esta tentación, será el resultado de su esfuerzo, pero si fracasa, será porque no *permitió* a su esfuerzo tener éxito. Y esto es debido al hecho de que, al tiempo que quería vencer la tentación, también quería fracasar, por razones totalmente diferentes y en competencia. Cuando los agentes, como la mujer, deciden en tales circunstancias, y los esfuerzos indeterminados que están haciendo se convierten en elecciones determinadas, *hacen* prevalecer un conjunto de razones o motivos en competencia sobre los otros, entonces y allí, *decidiendo* [...]. Sus actos de decisión, en otras palabras, son “will-setting”.

Kane propone que el conflicto interno que aparece en situaciones como esta se corresponde con un alejamiento del equilibrio termodinámico en regiones cerebrales apropiadas, de modo que las indeterminaciones que aparezcan a escala neuronal acceden al cerebro gracias a la situación de caos en la que este se encuentra. Así, “[l]o que se experimenta personalmente como incertidumbre corresponde físicamente a la apertura de una ventana de oportunidad que elimina temporalmente la completa determinación por influencias del pasado” (*id.*). Además, el esfuerzo de voluntad y el indeterminismo son inseparables: “[E]l esfuerzo *es* indeterminado y el indeterminismo es una *propiedad* del esfuerzo, no algo separado que ocurre después o antes del esfuerzo” (*ibid.*, p. 391).

En opinión de Kane, *la responsabilidad y el control* no tienen por qué verse empobrecidos por el indeterminismo. Imaginemos que cada una de las dos opciones motivacionales a las que se enfrenta la mujer de negocios viene representada por una red neural recurrente, y que ambas redes compiten y se influyen mutuamente. Cada opción se enfrenta al *ruido indeterminista* que supone que la alternativa esté presente. La decisión que tome la mujer viene representada por la llegada de una de las dos redes a un umbral de activación. En esta línea, Kane sostiene que “[c]ada tarea está siendo boicoteada por el indeterminismo creado por la presencia de la otra, así que podría fracasar. Pero si tiene éxito, entonces los agentes pueden ser considerados responsables porque [...] habrán tenido éxito en hacer lo que estaban intentando o esforzándose por hacer” (*ibid.*, p. 390). Así, el indeterminismo entorpece los procesos dirigidos a objetivos (“goal-directed”) que desembocan en las SFAs, por lo que no actúa como causa de una elección, pero sí goza de relevancia causal porque provoca que la probabilidad de que esta tenga lugar disminuya. Las razones para dicha elección, la consciencia acerca de estas y el esfuerzo en la deliberación sí poseen el rol de causas, puesto que provocan que aumente la probabilidad de que la elección tenga lugar.

Kane compatibiliza su postura con la visión de los agentes volitivos como sistemas dinámicos complejos que responden a la información. En su planteamiento, afirma, el todo y las partes se influyen entre sí, tal y como sucede en los sistemas dinámicos complejos: en el sistema motivacional al completo del agente, que él denomina “red del yo”, hay una amplificación del indeterminismo sináptico y neuronal que se topa con actividades dirigidas a objetivos que pertenecen a la red de escala superior. Asimismo, y tal y como ocurre en los sistemas complejos, hay capacidades emergentes: “Sólo cuando las criaturas logran el tipo de complejidad interna capaz de dar lugar a conflictos en sus voluntades, o sistemas motivacionales, surge la capacidad para la auto-formación característica del libre albedrío” (*ibid.*, p. 396).

El pensamiento de Kane ha recibido diversas críticas (consúltese, por ejemplo: Mele 1998, 1999; Moya 2006b, pp.151-62; Kane 2011, pp. 390-401). Por nuestra parte, nos gustaría señalar un problema de inteligibilidad que hallamos en los planteamientos del autor aquí expuestos, y que surge cuando nos preguntamos *cómo vence* una red neural recurrente a su competidora. Como ya hemos visto, las razones gozan de poder causal en la propuesta de Kane, pero *no son suficientes* para que tenga lugar una elección. Recordemos que, según Kane, las SFAs están indeterminadas porque no hay motivos o causas suficientes para que se den. Además, afirma que “el agente causa o provoca la elección que se hace involucrándose en un proceso dirigido a objetivos [consistente en] intentar o tratar de provocar esa elección (por *buenas* razones, aunque no *concluyentes* o *decisivas* [...]) y logrando alcanzar ese objetivo, cualquiera que sea la elección que se haga” (Kane 2011, p. 400). Entonces, si el poder causal de las razones no basta para que se produzca una elección, ¿qué factores participan para que así sea? Desde luego, no parece que el indeterminismo sea uno de ellos, pues, como vimos, Kane lo considera causalmente relevante pero no una causa. Y si no hubiera ninguno de dichos factores, ¿sería inteligible la aparición de un evento (en este caso, una elección) para el cual no existen causas suficientes? Responder afirmativamente constituiría un arriesgado

posicionamiento metafísico que requeriría una convincente justificación teórica. Por lo tanto, y a nuestro modo de ver, hay un importante vacío causal que socava la inteligibilidad de la relación entre razones y elecciones (SFAs) (cf. Mele 1998, pp. 582-3).

A pesar de sus elementos conflictivos, no cabe duda alguna de que la de Kane es la propuesta libertarista de índole materialista más elaborada e influyente que existe hasta la fecha, ni de que contiene elementos interesantes e inspiradores que pueden ser de enorme ayuda en la búsqueda de una mayor inteligibilidad del libre albedrío indeterminista. Por otro lado, se habrá observado que el indeterminismo ha sido concebido por Kane desde una perspectiva interna al agente. No sucede así con la propuesta que viene a continuación.

#### *4.3.2. Indeterminismo exógeno: la propuesta de Moya*

La propuesta de Carlos Moya (2006b, pp. 164-211 y 217-20) se caracteriza, entre otras cosas, por atribuir un origen externo a influencias indeterministas que permiten el control último por parte del agente. Aunque dicha propuesta está elaborada en relación a la responsabilidad moral, la consideraremos extensible al libre albedrío, respetando, así, un criterio que venimos siguiendo durante todo el trabajo: que la responsabilidad moral es consecuencia de nuestro libre albedrío. A continuación ofreceremos algunas ideas importantes para la comprensión del modo en que Moya concibe la génesis del indeterminismo y la relación de este con el control último.

El control último, sostiene Moya, es ininteligible en teorías que hacen un “*énfasis casi exclusivo en la voluntad y actos relacionados con la voluntad, especialmente elecciones*” (*ibid.*, p. 169), como la de Kane, por ejemplo. Pero el control último del agente no tiene por qué ser necesariamente voluntario o estar cimentado en sus elecciones. En este sentido, el autor sugiere trasladarnos de un enfoque

*individualista y conativo* a otro de carácter *social y cognitivo*. En otras palabras, no debemos poner el acento en nuestras elecciones o nuestros deseos, por ejemplo, sino en nuestras *creencias*, cuya relevancia destaca Moya: “La mayoría del tiempo actuamos sobre la base de nuestras creencias, de nuestra visión cognitiva de las cosas, que incluye nuestras creencias evaluativas sobre lo que es valioso y vale la pena perseguir, sin ninguna decisión consciente. Y en los casos en que la decisión está involucrada, cuando llega el instante de la decisión, gran parte del asunto ya está resuelto sobre la base de tales visiones cognitivas” (*ibid.*, p. 170).

Las creencias, que, según Moya, no poseen un carácter voluntario, pueden ser controladas de un modo involuntario pero no del todo pasivo. El control “no conlleva elección, sino más bien una actitud de humildad y respeto hacia los hechos. Elegir nuestras creencias es una manera de perder el control sobre ellas” (*ibid.*, p. 218). Pues bien, hay un tipo de creencias con una particular relevancia para la responsabilidad moral: aquellas que guardan un contenido con un carácter evaluativo respecto a lo que resulta trascendental perseguir en nuestras vidas y que podrían servir de referencia para nuestras elecciones y acciones. Moya establece seis condiciones, conectadas todas entre sí, que las creencias evaluativas deben cumplir para servir como base esencial de un eventual control último, y en consecuencia de una eventual responsabilidad moral, respecto a nuestras decisiones y nuestras acciones:

- 1) De acuerdo a la profundidad de las adscripciones de responsabilidad moral, deberían ser un componente central de una persona como agente potencialmente responsable moralmente.
- 2) Deberían ser correctamente atribuidas a un agente como su verdadero autor y origen, para que algunas de sus elecciones y acciones sean verdaderamente atribuibles a él como su fuente.
- 3) El agente debería tener control racional sobre esas creencias; deberían estar justificadas y basadas en razones. Este requisito, sin embargo, no debería dar lugar a una contraproducente regresión infinita.
- 4) Deberían ser potencialmente eficaces en nuestro comportamiento, aunque podamos a veces actuar contra ellas.
- 5) De acuerdo al aspecto regulativo del control último, deberíamos tener alternativas con respecto a ellas.
- 6) La condición de justificación (condición 3) debería sostenerse aunque las creencias no estén

causalmente determinadas; en otras palabras, nuestra propuesta no debería caer presa de alguna u otra versión del argumento de “Mind” (*ibid.*, p. 181).

Nos centraremos a continuación en la condición número 6, que hace referencia al indeterminismo. En opinión de Moya, este no permite el control racional si sigue un sentido ascendente desde el nivel subatómico a los niveles neurológico y mental. El autor califica este tipo de indeterminismo como “bottom-up”. Para que sea posible el control racional, y con él el libre albedrío y la responsabilidad moral, el indeterminismo “tiene que permanecer en los lugares apropiados y tener las relaciones apropiadas hacia los fenómenos mentales y neurológicos” (*ibid.*, pp. 195-6). Moya sugiere que se denomine “top-down” a esta clase de indeterminismo, y que este se encuentra sobre todo en *los sistemas, las prácticas y las instituciones de carácter normativo*.

Moya establece dos requisitos que su propuesta para la condición número 6 ha de cumplir. En primer lugar, es necesario demostrar que el indeterminismo se halla en los lugares mencionados y que no impide llevar a cabo un control racional de nuestras posibilidades alternativas. Podemos considerar que los sistemas, las prácticas y las instituciones de carácter normativo son indeterministas en tres sentidos distintos. Primero: es propiedad de dichos modelos que se diferencien las acciones que son o no son correctas en referencia a ellos.<sup>24</sup> Segundo: en muchas ocasiones ocurre que hay diversas acciones correctas y formas de satisfacer las exigencias de los patrones normativos o diversos actos incorrectos y maneras de incumplir dichas exigencias. Así, sucede a menudo que para decidir la adecuación de una acción a las exigencias de la pauta normativa cabe llevar a cabo una discusión basada en razones. Tercero: cabe una discusión basada en razones dirigida a si una circunstancia específica encaja en recibir la aplicación de un determinado juicio de índole normativa. Pues bien, en todos estos sentidos se encuentran indeterminados el *sistema de creencias* del agente (incluyendo las de carácter

---

<sup>24</sup> Moya cita aquí a Winch (1963).

evaluativo) y su *deliberación práctica* (con la que analiza posibles acciones y decide en consecuencia), sometidos a patrones de carácter normativo,<sup>25</sup> y, además, el indeterminismo resulta potenciado por la compleja interacción entre ambos. Todo ello sin menoscabo del control racional: “[E]ste indeterminismo relacionado con sistemas y prácticas normativas, a diferencia del indeterminismo relacionado con fenómenos cuánticos, está esencialmente vinculado a las razones y la discusión y la crítica razonadas” (*ibid.*, p. 198).

El segundo requisito establecido por Moya es el siguiente: es necesario explicar de modo inteligible cómo este indeterminismo recién caracterizado podría alcanzar nuestro nivel neurológico y configurarlo en parte. A este respecto, el autor sigue una línea de argumentación muy similar a la que describíamos en el apartado 4.2.1. (Moya 2011, pp. 197-200), cuya idea esencial, recordemos, era la siguiente: a lo largo del proceso de aprendizaje se refuerzan ciertas acciones del agente en virtud del contenido semántico de signos exógenos, produciéndose la selección y el moldeamiento de las conexiones nerviosas y neuromusculares que conducen a dichas acciones. Si bien los ejemplos utilizados por Moya tanto en aquel caso como en el trabajo que aquí analizábamos no tienen que ver con la responsabilidad moral de un modo directo, el autor piensa, creemos que con razón, que sus argumentos son extrapolables “a muchos otros sistemas y prácticas normativas, incluyendo aquellos directamente relacionados con la responsabilidad moral, tales como el sistema de creencias evaluativas de un agente y su deliberación práctica” (Moya 2006b, p. 202).

En nuestra opinión, Moya sigue una vía realmente interesante a la hora de apostar por el libertarismo sin abandonar el materialismo. La explicación ofrecida para el segundo requisito nos resulta especialmente inspiradora de cara al desarrollo de nuestra propia propuesta. De cualquier forma, para plantear esta seguiremos, a diferencia de Moya, una línea de tendencia individualista.

---

<sup>25</sup> Al hablar de la sujeción de la deliberación práctica a los modelos normativos, Moya cita a Hookway (2001, p. 190).

Ello no significa que pasemos por alto la dimensión social del libre albedrío, que también tendremos presente.

## Capítulo 5. Libre albedrío y neurociencia: una propuesta libertarista

«Para ciertos hombres... nada hay escrito si ellos no lo escriben.»

*Lawrence de Arabia* (David Lean, 1962).<sup>1</sup>

En el capítulo anterior hemos estudiado las posibilidades del libertarismo de hallar una explicación del libre albedrío en sintonía con la ciencia actual. Con este fin hemos analizado las que, en nuestra opinión, son las tres condiciones esenciales que han de satisfacerse para lograr dicho objetivo: causalidad mental, correlación inteligible entre la causalidad mental y un sustrato material, e indeterminación ontológica que permita el control último. Con estas condiciones como base fundamental, efectuaremos una propuesta que tratará de aliviar, al menos en parte, la dificultad que supone para el libertarismo la Cuestión de Inteligibilidad, planteada por Kane (1996, p. 13). Con la propuesta, que podemos enmarcar en la corriente del materialismo no reductivo,<sup>2</sup> no escondemos la intención de alcanzar una solución rotunda a dicha cuestión, ni mucho menos, sino que sugerimos una posible vía de investigación que consideramos prometedora y viable. Nos centraremos únicamente en la

---

<sup>1</sup> El guión de esta película es obra de Robert Bolt y Michael Wilson, y se basó en escritos autobiográficos de Thomas Edward Lawrence. El fragmento transcrito corresponde a la versión doblada al castellano.

<sup>2</sup> Véase sección 3.1. para un repaso panorámico de diversas posturas acerca del problema mente-cuerpo.

mencionada Cuestión de Inteligibilidad, y no en las otras cuatro cuestiones de Kane: de Compatibilidad, de Significatividad, de Existencia y Determinista. *La poda sináptica, los sistemas no integrables y los bucles extraños* constituyen elementos fundamentales de nuestra argumentación. Procederemos a continuación con el primero de dichos elementos, lo cual nos permitirá llegar a la primera parada en nuestro camino: una tesis que denominaremos “poda semántica”.

### 5.1. La poda sináptica

Como decíamos, un elemento clave de nuestra propuesta es el proceso de poda sináptica, o eliminación de sinapsis, conocido en neurociencia desde hace muchos años (*e.g.*, Huttenlocher 1979). Tapia y Lichtman (2008) han efectuado una descripción exhaustiva de este mecanismo neurobiológico, de la cual extraeremos a continuación algunas ideas fundamentales.

En esencia, la poda sináptica consiste en una doble disminución que tiene lugar durante el desarrollo del sistema nervioso: desciende la cantidad de axones que desembocan en una misma célula postsináptica y, por otra parte, lo hace también la cantidad de células postsinápticas en las que desemboca un mismo axón. Entre las conexiones afectadas por la poda se incluyen las neuromusculares, donde axones de tipo motor se unen con fibras musculares. Por otro lado, los autores sostienen que el mecanismo subyacente es la retracción de ramales de los axones (figura 5.2.). No obstante, debemos aclarar que tal mecanismo no es el único, pues también se ha observado la eliminación de espinas dendríticas (*e.g.*, Petanjek *et al.* 2011).<sup>3</sup> Sea como fuere, Tapia y Lichtman sostienen que, aunque los axones se desligan de algunas células

---

<sup>3</sup> Como información adicional, sirva apuntar que recientemente se ha observado en ratones que, en la fase de desarrollo postnatal, la microglía elimina sinapsis de una manera activa (Paolicelli *et al.* 2011).

postsinápticas, crece el número de sinapsis con aquellas otras a las que se mantienen vinculados, reforzándose así las conexiones sinápticas supervivientes (figura 5.1). De esta forma, y en virtud de una competencia entre neuronas, tiene lugar una reconfiguración de la red de transmisiones sinápticas en la que se pasa de un alto grado de redundancia y solapamiento a la aparición de circuitos únicos. Esto podría constituir una adaptación del individuo a su ambiente concreto, con posible fundamento en experiencias que gozan de influencia sobre el desarrollo del sistema nervioso.

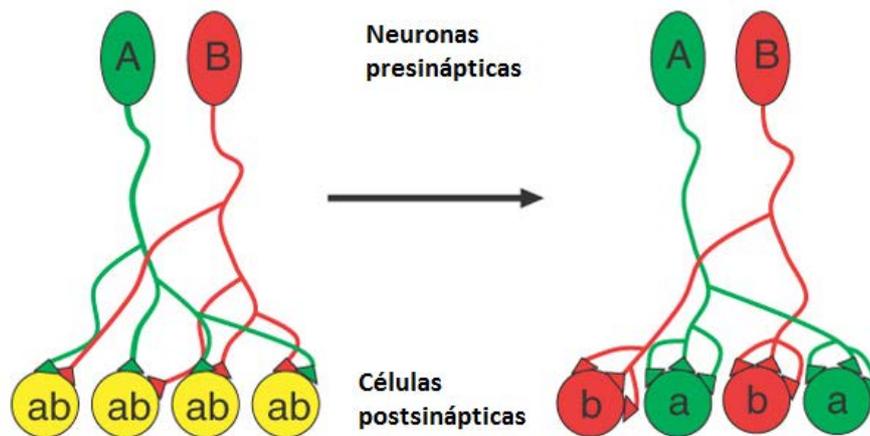


Figura 5.1. Durante el proceso de poda sináptica: 1) disminuye el número de axones que van a parar a una misma célula postsináptica y también el número de células de este tipo a las que va a parar un mismo axón, y 2) se refuerzan los vínculos intercelulares supervivientes mediante la formación de sinapsis adicionales. Adaptado de Tapia y Lichtman (2008).

Tapia y Lichtman concluyen su trabajo preguntándose si la poda neuronal ha de relacionarse en exclusiva con el proceso de desarrollo. A este respecto, escriben:

El tipo más importante de plasticidad adulta debe ser sin duda la memoria. ¿Podría tener la eliminación de sinapsis algo que ver con la memoria? Varios neurobiólogos [...] han dado explícitamente argumentos a favor de que la selección (en lugar de la instrucción) juega potencialmente un importante

papel en el aprendizaje. La idea es que, en el cerebro, el sistema de circuitos sinápticos existe *a priori* para muchas cosas que puede que finalmente se aprendan, de manera que el aprendizaje podría ocurrir mediante la selección de caminos sinápticos que ya existen en lugar de la construcción de nuevos circuitos. Aunque dicha selección podría ocurrir incrementando la fortaleza de un conjunto de interconexiones sinápticas o mediante el debilitamiento de otros, también podría ocurrir eliminando completamente algunos circuitos. Es importante resaltar la distinción entre la plasticidad que altera las fortalezas de conexiones existentes y el tipo de plasticidad más extremo [...] que provoca la erradicación permanente del *input* [cursiva añadida] de un axón hacia células postsinápticas particulares (Tapia y Lichtman 2008, p. 488).

La eliminación de *inputs*, añaden los autores, podría servir para aislar a un circuito sináptico del ruido causado por *inputs* transmitidos a destiempo, evitando así que unos contenidos sean borrados al memorizarse otros nuevos (*id.*). Pero explicar el mecanismo memorístico mediante una poda sináptica como la descrita necesitaría de la eliminación de sinapsis también en el cerebro de un adulto (*id.*). A raíz de trabajos como el de Huttenlocher (1979), cabría quizá pensar que esta eliminación no es posible, dado que este observó que a los dos años de vida comenzaba un descenso en la densidad de las sinapsis del córtex frontal del cerebro humano, y que dicho descenso finalizaba a los dieciséis, edad a partir de la cual se mantenía estable hasta la vejez, cuando tenía lugar un pequeño descenso adicional. Sin embargo, pensamos que existen, al menos, cuatro motivos razonables para pensar que la eliminación de sinapsis, una vez desencadenada, puede quizá perpetuarse a lo largo de toda la vida:

(1<sup>o</sup>) En un trabajo reciente, Petanjek *et al.* (2011) han observado que la poda sináptica de espinas dendríticas en el córtex prefrontal de nuestra especie sobrepasa el período adolescente y sigue operando durante la fase entre los veinte y los treinta años, tras lo cual tiene lugar una estabilización. Con este hallazgo queda considerablemente ampliado el límite temporal de la eliminación de sinapsis, que quedaba establecido hasta ahora en la adolescencia, y se llega a edades correspondientes a individuos ya adultos. No sería descabellado pensar que en un futuro se puedan efectuar nuevos

descubrimientos, quizá en otras regiones neurales (léase el párrafo siguiente), que extiendan nuevamente el límite. Además, y aunque los autores no extraen conclusiones de ello, en las curvas de regresión trazadas en el artículo (*ibid.*, p. 13283) puede observarse que la estabilización que sigue a los treinta años no viene representada por una línea horizontal sino por un descenso en la densidad de espinas dendríticas que, si bien es suave y cada vez más paulatino, no es menos cierto que continua hasta la vejez (figura 5.2.). A nuestro modo de ver, todo ello deja abierta la posibilidad de que la poda sináptica sea un proceso que, conforme a las necesidades propias de cada fase vital, muestre diferentes intensidades a lo largo de una vida humana completa.

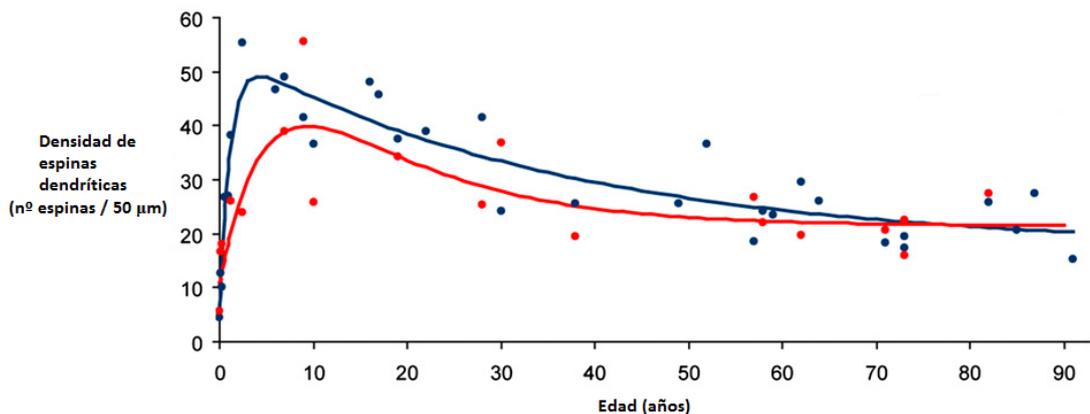


Figura 5.2. Dos de las curvas de regresión mencionadas en el párrafo anterior. Los dos colores corresponden a dos capas neuronales distintas del córtex prefrontal dorsolateral humano. Adaptado de Petanjek *et al.* (2011).

(2º) Gracias al trabajo de, nuevamente, Petanjek *et al.* y a otros anteriores (*e.g.*, Huttenlocher y Dabholkar 1997) ha quedado de manifiesto que el patrón temporal de la eliminación de sinapsis difiere entre distintas regiones del córtex cerebral, entre capas de la misma región cortical e, incluso, entre subcapas compuestas por distintas clases de dendritas que contienen las neuronas de una misma capa. No sería descartable que existan, aún por

descubrir, regiones, capas o hasta subcapas donde la poda se dé continuamente.

(3<sup>o</sup>) Anteriormente apuntamos que durante el proceso de poda sináptica se refuerzan las conexiones supervivientes al crecer el número de sinapsis que los axones trazan con las células postsinápticas a las que se mantienen vinculados (ver figura 5.1.). Por tanto, cabe la opción de que la estabilización de la densidad sináptica en el tiempo se deba, al menos en parte, a esta reorganización. Dicho de otro modo: durante la trayectoria vital correspondiente a la estabilización, la eliminación de algunos vínculos intercelulares se vería cuantitativamente compensada en gran medida por el robustecimiento de otros y, por tanto, la poda sináptica, si bien en menor grado que en el periodo previo, estaría teniendo lugar aún.

(4<sup>o</sup>) El hecho de que, tras un marcado descenso en la densidad sináptica, aparezca una estabilización no indica necesariamente que la eliminación de sinapsis finalice. Spalding *et al.* (2013) han observado la producción diaria de unas 1400 neuronas nuevas en el hipocampo humano adulto, y han propuesto que esta neurogénesis podría ser importante en el cerebro desde el punto de vista funcional. Debería considerarse la posibilidad de una neurogénesis cuantitativamente relevante en otras áreas de nuestro sistema nervioso, o incluso una neurogénesis ubicua, que sirva como contrapeso a una posible poda sináptica en los individuos en plena fase adulta. Desde este punto de vista, la estabilización en la densidad sináptica observada a partir de los treinta años podría explicarse, cuando menos parcialmente, por un equilibrio entre la posible pérdida adulta de sinapsis y la aparición de nuevas conexiones sinápticas en virtud de una neurogénesis más o menos generalizada.<sup>4</sup>

---

<sup>4</sup> Una argumentación similar a esta, pero referente a la fase de desarrollo del sistema nervioso central, puede leerse en Tapia y Lichtman (2008, p. 469).

Terminamos aquí nuestro tratamiento del mecanismo de poda sináptica. Veamos ahora cuál es el rol que desempeña en nuestros planteamientos.

## 5.2. La “poda semántica” como modalidad de causalidad mental

En el capítulo anterior, finalizábamos la sección 4.2. formalizando nuestra concepción de causalidad mental de la siguiente manera:

Tesis  $\gamma$  (CAUSALIDAD MENTAL).

La imposición de determinadas constricciones de carácter semántico por parte del contexto normativo en la mente de un ser humano (la causa) hace que el proceso de selección y refuerzo de las rutas neurales y neuromusculares conducentes a la ejecución de ciertas acciones (el efecto) sea más probable que en ausencia de esas constricciones, siendo indispensable que dicho proceso finalmente se produzca.

Consideramos que la poda sináptica podría ser buena candidata como mecanismo neurobiológico interviniente en el proceso de selección y refuerzo de rutas neurales y neuromusculares mencionado en esta tesis. Se trataría, por tanto, de “el efecto” en  $\gamma$ , o, quizá, de un tipo de efecto entre varios posibles. La compatibilidad de la poda sináptica con  $\gamma$  implicaría, por supuesto, su compatibilidad con la emergencia contextual, con el papel de esta en el ser humano entendido como sistema complejo y con la propuesta de Moya para armonizar materialismo y causalidad mental (o, al menos, con los aspectos

esenciales de la misma). Con semejante enfoque ha sido concebida la aproximación que a continuación sugerimos:<sup>5</sup>

[I] *Poda semántica*. A lo largo de un proceso de aprendizaje consistente en la adquisición de habilidades sujetas a reglas de carácter semántico, tales como el lenguaje ordinario o la interpretación musical, la adecuación de ciertos resultados en forma de sonidos, hablas o acciones corporales al estándar establecido por las reglas semánticas constituye una constricción contextual de índole social que provoca la selección de las rutas neurales y neuromusculares que conducen a la producción de dichos resultados. Esta selección constituye el inicio de un mecanismo de poda sináptica que, con la práctica continuada y el perfeccionamiento de las habilidades mencionadas, actúa de forma paulatina y sostenida en el tiempo, logrando un refuerzo creciente de las rutas seleccionadas en un principio. Todo este proceso, al que denominaremos *poda semántica*, implica un vínculo de emergencia contextual en el que la descripción de la configuración neural y neuromuscular de un individuo humano, que es un sistema dinámico complejo, resulta necesaria pero insuficiente para completar la explicación del proceso de aprendizaje, para lo cual debe acudir al estándar semántico establecido por el contexto social.

[II] *Desarrollo del proceso*. El refuerzo efectuado en la poda semántica progresa de dos maneras complementarias. Por un lado, los enlaces sinápticos situados en las rutas neurales y neuromusculares que, dada su inadecuación al estándar semántico, son descartadas (no seleccionadas) por constricción contextual, van desapareciendo mediante la retracción de ramales de los axones y/o la eliminación de espinas dendríticas, produciéndose una rebaja progresiva del número de axones que desembocan en una misma célula postsináptica y de la cantidad de células postsinápticas con las que conecta un mismo axón. Por otro lado, en cambio, tiene lugar la formación de sinapsis adicionales entre las

---

<sup>5</sup> Percibirá el lector que el tiempo verbal mayoritariamente empleado en la redacción de nuestra propuesta ha sido el presente de indicativo. Con ello no pretendemos mostrarnos categóricos, sino hacer más ágil la lectura.

células que pertenecen a las rutas que han sido seleccionadas al iniciarse el aprendizaje.

[III] *Convergencia entre individuos.* De esta manera, la estructura única que, más allá de las características comunes a los miembros de una misma especie, muestra cada ser humano en su red neural y neuromuscular se reconfigura con el transcurso del tiempo para conformar circuitos con redundancia y solapamiento gradualmente menores, y adaptados al ambiente social donde se desenvuelve el individuo. Esto constituye una suerte de “convergencia adaptativa” que, de un modo análogo a, por ejemplo, el proceso de selección natural que dio lugar a la formación de las alas en las abejas, los murciélagos y las águilas, crea patrones similares (aunque no idénticos) en individuos con estructuras neurales y neuromusculares inicialmente dispares, en virtud de una competencia interneuronal. Este sería el caso de los violinistas perfectamente sincronizados en el ejemplo planteado por Moya.

[IV] *Relación con la memoria.* La reconfiguración neural y neuromuscular progresivamente efectuada por la poda semántica marca una especie de impronta memorística, que permite una ejecución cada vez más eficaz de los sonidos, hablas o acciones corporales que se emplearán en el desempeño de las habilidades aprendidas y, por qué no, en el de otras con similares características que puedan sobrevenir (cf. Moya 2006b, pp. 199-200). En otras palabras: a través de la poda sináptica, el aprendizaje de esas habilidades queda “fijado” en los tejidos biológicos. Se consigue, así, que un uso de la memoria basado, en esencia, en la memoria de trabajo (*e.g.*: un aprendiz de violinista observa a su maestro tocar un nuevo acorde e inmediatamente trata de reproducirlo) evolucione a un empleo de la memoria a largo plazo principalmente (*e.g.*: el aprendiz, convertido ya en violinista de la orquesta de Moya, ejecuta de un modo rápido y cuasi automático el primer movimiento de la Quinta Sinfonía de Beethoven).

[V] *Balance entre neurogénesis y poda sináptica.* El número total de sinapsis en el sistema nervioso de un individuo cambia a lo largo de su vida como

resultado de variaciones en el balance entre la creación de nuevas conexiones sinápticas (por medio de la neurogénesis) y la destrucción de otras ya existentes (por medio de la poda sináptica). Diremos que hay un “balance sináptico” positivo cuando la creación supere a la destrucción, y negativo en caso contrario. Atendiendo al signo de dicho balance podemos establecer, *grasso modo*, tres grandes etapas vitales postnatales. En la primera de ellas, que transcurre entre el nacimiento y los dos años aproximadamente, el balance es positivo, debido principalmente a la creación frenética de nuevas conexiones sinápticas en un sistema nervioso aún incompletamente desarrollado. Una causa adicional podría ser una escasa eliminación de sinapsis como resultado, primero, de la relativamente baja dependencia del niño respecto a su ambiente más allá de su entorno más cercano (familia y hogar), y, segundo, de su limitada capacidad para desarrollar habilidades sujetas a reglas de carácter semántico. La segunda etapa, con inicio a los dos años y final en unos treinta (aunque no es descartable que se prolongue más allá), muestra balance negativo, por causa de una disminución en la intensidad de la neurogénesis y un incremento de la poda sináptica. La neurogénesis disminuye porque el desarrollo del sistema nervioso está muy avanzado y, en consecuencia, no se requiere una formación ingente de nuevas neuronas y sinapsis. En cuanto al incremento en la eliminación de sinapsis, nace de dos motivos: existe la necesidad de limitar el número de rutas neurales y neuromusculares para lograr una adaptación más eficaz al ambiente; pero, además, hay un aprendizaje enormemente significativo de habilidades sujetas a estándares semánticos, como el lenguaje hablado y escrito u otros lenguajes de tipo visual (*e.g.*: las señales de tráfico), sonoro (*e.g.*: la música, la sirena de ambulancia) o incluso táctil (*e.g.*: el braille, los besos), por lo que tiene lugar una poda semántica de considerable magnitud. La tercera etapa, finalmente, llega hasta el fallecimiento y se caracteriza por un balance sináptico nulo o levemente negativo, resultante de un equilibrio o cuasi-equilibrio entre la neurogénesis y la eliminación de sinapsis. Ambos procesos operarán con bajas tasas de incidencia. La neurogénesis ya había frenado su

ritmo durante la etapa previa; además, el sistema nervioso goza ahora de pleno desarrollo. Por lo que respecta a la poda de sinapsis, la adaptación ambiental es prácticamente total y ya no resulta imprescindible reducir el número de vías neurales y neuromusculares para ese fin. Además, en la edad adulta plena y en la vejez se reducirá el aprendizaje de habilidades sujetas a reglas de índole semántica, por lo que la poda semántica actuará en menor medida. Sin embargo, este último mecanismo, en conjunción con la neurogénesis adulta, seguirá operando y teniendo su importancia desde el punto de vista funcional: la formación de nuevas neuronas abre nuevas rutas, inyectando así nueva variabilidad neural, a partir de la cual la poda semántica moldeará circuitos únicos asociados a nuevos procesos de aprendizaje. Así, el ser humano consigue desarrollar destrezas relacionadas con sistemas semánticos normativos en todas las épocas de su vida.

[VI] *Matizaciones.* Deben tenerse en cuenta algunos factores que pueden introducir variantes en la secuenciación esquemática de fases vitales que acabamos de exponer. Por ejemplo, la poda sináptica exhibe distintas magnitudes y patrones temporales según la región, capa o subcapa del sistema nervioso donde tenga lugar. También son importantes la variabilidad genética y la ambiental, de tal forma que individuos con distinta composición génica (incluyendo, por ejemplo, la relativa a la raza y al sexo) o los que habiten en lugares con disparidad de características climáticas, sociales, culturales, etc., mostrarán balances sinápticos desiguales y etapas vitales con diferentes duraciones.

[VII] *Tesis final.* En consecuencia con todo lo anterior, proponemos emprender una vía de investigación en la que se establezca un vínculo concreto entre el proceso neurobiológico de eliminación (poda) de sinapsis y el concepto de causalidad mental establecido en la tesis  $\gamma$ . Recordemos dicha tesis (que fue presentada en el apartado 4.2.3.):

#### Tesis $\gamma$ (CAUSALIDAD MENTAL).

La imposición de determinadas constricciones de carácter semántico por parte del contexto normativo en la mente de un ser humano (la causa) hace que el proceso de selección y refuerzo de las rutas neurales y neuromusculares conducentes a la ejecución de ciertas acciones (el efecto) sea más probable que en ausencia de esas constricciones, siendo indispensable que dicho proceso finalmente se produzca.

El vínculo entre la poda sináptica y este concepto de causalidad mental consistiría en la consideración de la poda sináptica como el efecto causal en  $\gamma$ , lo que nos permite acotar dicha tesis al caso específico de una posible modalidad de causalidad mental que nos hemos dado en llamar *poda semántica*.

#### Tesis $\delta$ (PODA SEMÁNTICA).

La imposición de determinadas constricciones de carácter semántico por parte del contexto normativo social en la mente de un ser humano (la causa) hace que el proceso de poda sináptica en el que se seleccionan y refuerzan las rutas neurales y neuromusculares conducentes a la ejecución de ciertas acciones (el efecto) sea más probable que en ausencia de esas constricciones, siendo indispensable que dicho proceso finalmente se produzca.

Tal y como sucedía con  $\gamma$ , cabría considerar esta tesis como una versión de la propuesta dretskeana (a través de la de Moya): determinadas conexiones neurales y neuromusculares son *reclutadas*, mediante poda sináptica, para ejercer funciones concretas debido a aquello que *indican* sobre los signos presentes en el contexto normativo.

Con  $\delta$  pretendemos aproximarnos a las dos primeras condiciones esenciales que nos venimos marcando, ya desde el capítulo 4, para dar con una teoría libertarista y materialista del libre albedrío: la causalidad mental y la

correlación inteligible entre esta y un sustrato material. Es hora de preguntarnos por la tercera: la indeterminación ontológica que permita el control último.

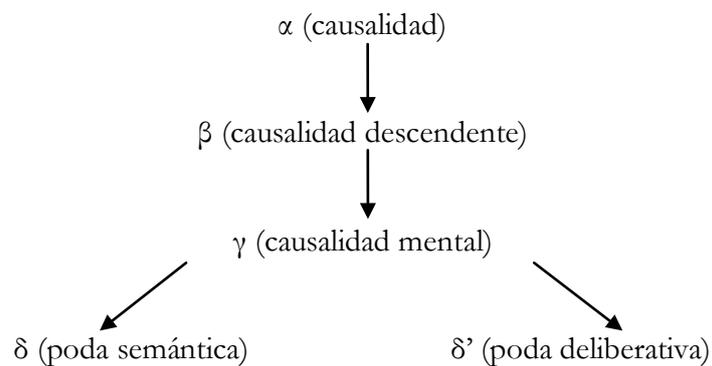
### **5.3. Indeterminismo y causalidad mental: las vías exógena y endógena**

Bajo nuestro punto de vista, el concepto de poda semántica es compatible casi a la perfección con la propuesta defendida por Moya. A su juicio, recordemos (véase apartado 4.2.1.), en el progreso de las habilidades del habla o de la interpretación musical hay un lugar para la causalidad descendente por medio de la selección y el refuerzo de determinadas conexiones neurales y neuromusculares en virtud del contenido y el significado de los signos lingüísticos y musicales, respectivamente. La única discordancia reseñable entre nuestra postura y la de Moya nace del hecho de que, en su opinión, el desarrollo de las destrezas del habla o de la interpretación musical necesita de un crecimiento en la complejidad de las conexiones neurales (Moya 2011, p. 197; cf. Moya 2006b, p. 200), mientras que la poda semántica traza una vía opuesta en este aspecto: la red de conexiones neurales y neuromusculares se simplifica por medio de un proceso de poda sináptica. Por otro lado, y tal y como sucedía con  $\gamma$  (véase apartado 4.2.3.), cabría considerar nuestra tesis final como una versión de la propuesta dretskeana (a través de la de Moya): determinadas conexiones neurales y neuromusculares son *reclutadas*, mediante poda sináptica, para ejercer funciones concretas debido a aquello que *indican* sobre los signos presentes en el contexto normativo social.

Pensamos también que la poda semántica es compatible con el indeterminismo de tipo exógeno concebido por Moya como elemento que hace posible el control racional del agente y, con él, el libre albedrío y la responsabilidad moral. Como ya explicamos en el apartado 4.3.2., el autor

sostiene que este indeterminismo se halla, sobre todo, en los sistemas, las prácticas y las instituciones de carácter normativo. La compatibilidad con  $\delta$  se dará si entendemos estos tres componentes como elementos contextuales que imponen constricciones semánticas en la mente de un ser humano. De esta manera, la poda semántica podría ser planteada como hipótesis materialista y libertarista del libre albedrío, pues satisfaría, en principio, las tres condiciones fundamentales que al respecto venimos proponiendo. De cualquier forma, y como ya adelantábamos en el capítulo previo, trazaremos una vía individualista, diferente a la seguida por Moya en la búsqueda de un indeterminismo que permita el control último. En otras palabras, seguiremos una vía endógena. Sin embargo, no aportaremos argumentos ni a favor ni en contra del indeterminismo propuesto por Moya, aunque, como el propio lector podrá comprobar, nuestra alternativa no excluye dicha posibilidad ni su convivencia con el indeterminismo endógeno que defenderemos.

En  $\delta$ , las constricciones normativas que desencadenan el proceso de poda sináptica son de carácter ambiental. La opción alternativa que propondremos estará centrada en constricciones normativas de carácter interno, recibirá el nombre de *poda deliberativa* y quedará formalizada mediante la tesis  $\delta'$ . Las relaciones entre las tesis que venimos planteando desde el capítulo anterior quedan de la siguiente manera:



Trazada ya la vía exógena por medio de  $\delta$ , seguiremos ahora la vía endógena. Comenzaremos hablando sobre los sistemas no integrables y los bucles extraños, pues ambos constituyen elementos esenciales para el planteamiento de  $\delta'$  (el tercero es la poda sináptica).

#### **5.4. El Grupo Bruselas-Austin y los sistemas alejados del equilibrio termodinámico**

Dado que venimos asumiendo que los sistemas complejos son el sustrato material adecuado como correlato de la causalidad mental, es necesario que nos preguntemos si es posible encontrar en ellos mecanismos ontológicamente indeterminados que posibiliten el control último. A este respecto, resulta del todo interesante la investigación que el grupo liderado por Ilya Prigogine ha llevado a cabo acerca de los sistemas alejados del equilibrio termodinámico (cf. Bishop 2011b, p. 124). A continuación, y basándonos en varios trabajos importantes de este equipo (Antonioniou y Prigogine 1993, Petrosky y Prigogine 1996 y 1997, Prigogine 1997, Prigogine 2003), conocido como “Grupo Bruselas-Austin”, trataremos sobre los aspectos más destacables de su investigación y sobre las principales conclusiones a las que han llegado.

La descripción de los sistemas dinámicos estudiados en la física clásica se realiza por medio de la función hamiltoniana, es decir, “la expresión de la energía en términos de las observables  $p$  y  $q$ ” (Prigogine 2003, p. 11), siendo  $p$  los momentos y  $q$  las coordenadas de los elementos del sistema. Así sucede, por ejemplo, con el estudio de las partículas de un gas. La función hamiltoniana nos permite adelantar cómo se desarrollará el movimiento. No obstante, está comúnmente aceptada la eliminación de las coordenadas, de manera que la función dependerá exclusivamente de  $p$ . Ahora bien, gracias a Poincaré, sabemos desde el siglo XIX que dicha práctica puede llevarse a cabo

únicamente con ciertos sistemas dinámicos: los conocidos como *sistemas integrables*. Imaginemos el caso de un gas compuesto por un gran número de partículas: la eliminación de las coordenadas conllevará una caracterización según la cual el movimiento de una partícula es concebido de forma independiente al resto. En un sistema integrable se trata con “objetos *independientes*, ya que las interacciones son eliminadas, o mejor dicho, incluidas en la definición de estos objetos” (*ibid.*, p. 12).

Al contrario de lo que ocurre en los sistemas integrables, hay ocasiones en las que no es posible suprimir las interacciones. Tal es la situación en los *sistemas no integrables*, que, lejos de ser inusuales, constituyen la norma en el mundo natural. La descripción de estos sistemas se realiza a través de distribuciones, no por medio de puntos. Dichas distribuciones presentan irreductibilidad a las trayectorias individuales, y su estudio es de carácter estadístico:

Es bien conocido, desde el trabajo pionero de Gibbs y Einstein, que podemos describir la dinámica desde dos puntos de vista. Por un lado, tenemos la descripción *individual* en términos de trayectorias (o de funciones de onda); por el otro lado, la descripción estadística en términos de distribuciones de probabilidad [...]. Siempre se asumió que los dos niveles de descripción eran equivalentes. [...] [E]sto no es así para las clases de sistemas donde esperamos que surjan procesos irreversibles. Entonces la descripción básica es en términos de la descripción estadística. Obtenemos nuevas soluciones para la distribución de probabilidad que son “irreducibles” puesto que no se aplican a las trayectorias individuales (Petrosky y Prigogine 1997, pp. 1-2).

Tenemos, así, posibilidades en lugar de certezas. También irreversibilidad, que nace del hecho de que se quebranta la simetría entre el tiempo pasado y el tiempo futuro, lo cual marca una *flecha del tiempo*. Pero la flecha no es puramente fenomenológica: la irreversibilidad no tiene que ver con nuestro conocimiento incompleto del sistema, sino que es intrínseca a este. Además, provoca efectos colectivos en los que interviene una infinidad de partículas que actúan de modo coherente. Las partículas mantienen entre sí interacciones de carácter persistente, no transitorio, por lo que el sistema debe ser concebido de

un modo holístico, no local. En este sentido, Prigogine (2003, p. 18) afirma: “[L]as interacciones son una parte fundamental de la naturaleza que observamos y, en los sistemas no integrables, las interacciones no pueden ser eliminadas. [...] En un gas, incluso si está en equilibrio, las colisiones continúan ocurriendo y las interacciones nunca son eliminadas. Las colisiones dan lugar al movimiento térmico. Hay límites para el reduccionismo”. La irreversibilidad es una propiedad emergente en la medida en que no es posible su comprensión desde un punto de vista local y basado en interacciones transitorias.

En definitiva, y tal como destaca Robert Bishop (2002, p. 120; 2011a, p. 92), los sistemas no integrables se caracterizan por poseer o mostrar: propiedades de carácter emergente, numerosas partículas, actuación colectiva, estructura y orden considerables, y evolución irreversible. Existe la posibilidad de emplear las investigaciones de Prigogine y su grupo para el estudio del cerebro concebido como sistema alejado del equilibrio termodinámico, ya que este órgano presenta los cinco atributos mencionados (*id.*). El propio Bishop ha mostrado un gran interés tanto en las investigaciones del Grupo Bruselas-Austin como en su repercusión para el problema del libre albedrío (véanse: Bishop 2002, pp. 120-21; 2004; 2011a, pp. 92-3; 2011b, pp. 124-5). Sobre las implicaciones de dichas investigaciones para el indeterminismo, sostiene que

centrarse exclusivamente en las funciones de distribución abre la posibilidad de que los modelos de no equilibrio macroscópicos sean irreductiblemente indeterministas, un indeterminismo que no tiene nada que ver con la ignorancia acerca del sistema o los caprichos de la mecánica cuántica. De ser cierto, esto significaría que las probabilidades son un elemento tan ontológicamente fundamental del mundo macroscópico como lo son del mundo microscópico, estando a la vez libres de las dificultades interpretativas encontradas en la mecánica cuántica convencional (Bishop 2011a, p. 93).<sup>6</sup>

---

<sup>6</sup> Recordemos que en el capítulo 2 del presente trabajo ya se trató sobre el problema indeterminismo *versus* determinismo en la mecánica cuántica (consúltese especialmente el apartado 2.6.4.).

El autor aclara que quedan pendientes de solución problemas importantes, como el modo adecuado de interpretar las distribuciones de probabilidad, o como el origen físico del indeterminismo mencionado (Bishop 2011b, p. 124). Sobre el segundo asunto, sugiere una opción: el indeterminismo podría emerger, desde el punto de vista ontológico, a partir del comportamiento dinámico que se produce a nivel inferior (*id.*). Es posible que se refiera a esto cuando propone que “las [...] trayectorias del sistema dan las condiciones necesarias para la existencia de la distribución [...], pero *no* condiciones *suficientes* para su evolución” (Bishop 2004, pp. 25-6). Nótese que Bishop está apelando (aunque no utiliza el término) a un vínculo de *emergencia contextual* entre las trayectorias individuales y la distribución de probabilidad. Nosotros asumiremos dicho vínculo y, partiendo de la consideración del cerebro humano como sistema no integrable, propondremos que, en ciertas ocasiones, para completar la explicación de algunas modificaciones estructurales sufridas por este órgano es necesario incluir constricciones de tipo interno impuestas por un patrón abstracto con forma de bucle extraño. Esta será la vía para plantear la tesis de la poda deliberativa (tesis  $\delta$ ).<sup>7</sup> Veamos ahora qué son los bucles extraños y qué clase de contribuciones pueden aportar al problema de las relaciones entre cuerpo y mente.

## 5.5. Hofstadter y los bucles extraños

Hace casi cuatro décadas, Douglas Hofstadter (1979) publicó su célebre obra *Gödel, Escher, Bach*, en la que estudia a fondo tanto el concepto de bucle extraño

---

<sup>7</sup> Estrictamente hablando, el símbolo  $\delta$  alude a la descripción formalizada de la poda deliberativa como relación causal. No obstante, y con el objetivo de aligerar la lectura, de ahora en adelante utilizaremos frecuentemente el símbolo para referirnos a la tesis completa de la poda deliberativa. Hacemos extensible este criterio a la poda semántica y el símbolo  $\delta$ .

como su relación con la naturaleza de la conciencia y el problema mente-cerebro. Recientemente, este autor ha seguido la misma línea en *I Am a Strange Loop* (Hofstadter 2007). En estos libros defiende varias ideas de gran relevancia para el contenido de este capítulo, y sobre ellas hablaremos en esta sección.



Figura 5.3. Fotogramas correspondientes a tres de los patrones de realimentación de vídeo obtenidos por Hofstadter. Tomados de Hofstadter (2007).

Pensemos en el siguiente experimento. Disponemos de un aparato de televisión que emite una determinada señal. Además, colocamos una cámara de vídeo que graba las imágenes que está reproduciendo la pantalla del televisor. Sin embargo, dichas imágenes corresponden, precisamente, a aquello que la cámara está grabando, que no es otra cosa que la emisión de la propia televisión. Esta situación provoca un bucle (ciclo) de realimentación de vídeo. Hofstadter realizó este experimento y obtuvo diversos patrones dependiendo

de los ángulos horizontal y vertical de la cámara, de la configuración de diferentes funciones de esta (contraste, zoom, etc.) y de la colocación de objetos entre el objetivo y la pantalla. Algunos eran muy sencillos y mostraban un comportamiento estático (figura 5.3). Otros, en cambio, resultaron mucho más complejos. A veces, esto sucedía al situar cuerpos (como las manos) delante de la cámara (figura 5.4).

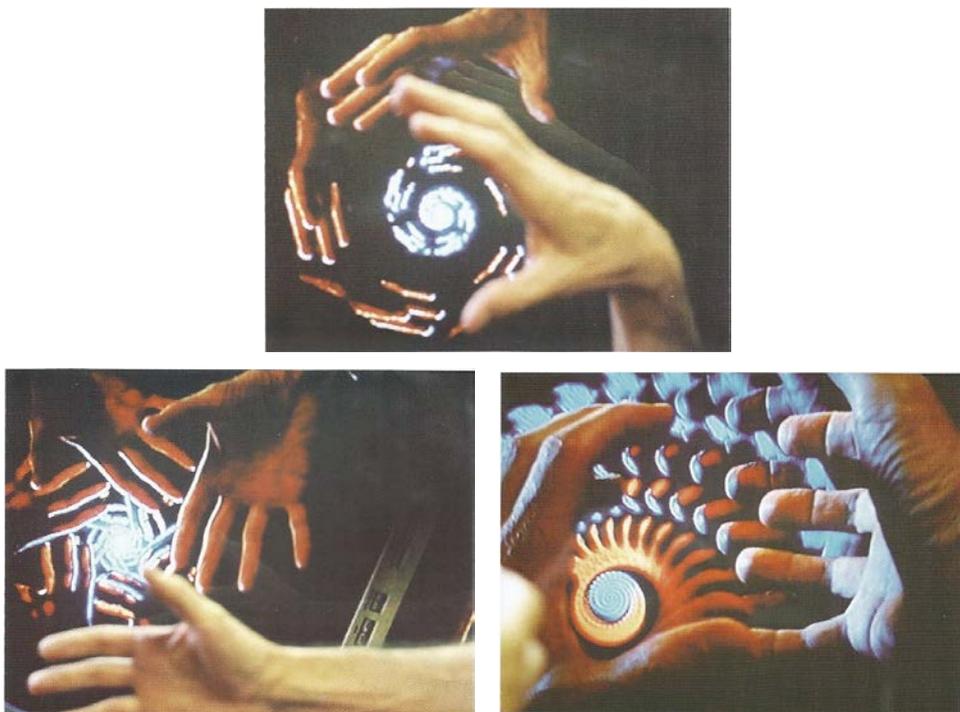


Figura 5.4. Fotogramas correspondientes a patrones dinámicos con forma de espiral obtenidos por Hofstadter. Tomados de Hofstadter (2007).

Hofstadter interpreta los patrones de su experimento como fenómenos que emergen a partir del nivel inferior a ellos y que se mantienen a sí mismos indefinidamente. Así,

la realimentación da lugar a un nuevo tipo de fenómeno abstracto que puede denominarse “bloqueo”. Justo a partir de la más débil señal (la primerísima

imagen enviada a la pantalla de TV en la primera diminuta fracción de segundo), aparece, casi al instante [...], la realización completa de todas las implicaciones de esa señal – y esa nueva estructura de nivel superior, ese patrón emergente en la pantalla, ese epifenómeno, está entonces “bloqueado” gracias al bucle. No desaparecerá porque está siempre actualizándose a sí mismo, alimentándose a sí mismo, renovándose a sí mismo. Dicho de otro modo, el patrón emergente resultante es una estructura auto-estabilizadora cuyos orígenes, a pesar de la simplicidad del bucle de realimentación en sí, son casi impenetrables porque el bucle realiza el ciclo en una gran cantidad de ocasiones (Hofstadter 2007, p. 70).

Hofstadter utiliza este fenómeno (y así lo reconoce) como metáfora para defender sus argumentos, y, como se verá más adelante, tomaremos prestada dicha metáfora para nuestros propios razonamientos. Ahora bien, el autor aclara que los bucles surgidos de la realimentación de vídeo no son bucles extraños. ¿Qué es, entonces, un bucle extraño?

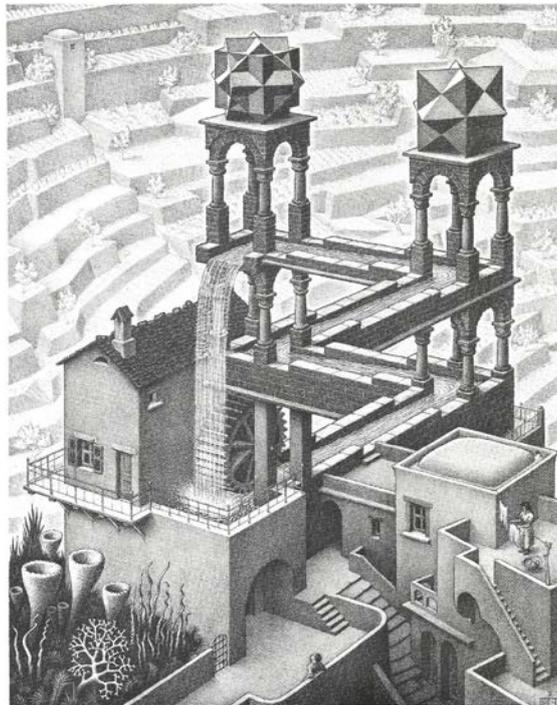


Figura 5.5. *Waterfall*, de M. C. Escher. Tomado de Escher (1992).

El artista M. C. Escher es el autor de varias obras que Hofstadter emplea para ilustrar el concepto de bucle extraño.<sup>8</sup> Un ejemplo es *Waterfall* (figura 5.5), donde puede apreciarse una cascada cuya agua desciende continuamente y en varios tramos de un modo tal que siempre recorre el mismo trayecto. Estamos ante un bucle extraño: un fenómeno, dice Hofstadter, según el cual un ascenso o un descenso por una sucesión de niveles jerárquicos acaba conduciendo al nivel desde el cual se comenzó. Se trata

no [de] un circuito físico sino [de] un bucle *abstracto* [la cursiva es nuestra] en el que, en la sucesión de etapas que constituyen la vuelta al ciclo, hay un cambio de un nivel de abstracción (o estructura) a otro, lo cual parece un movimiento hacia arriba en una jerarquía, y sin embargo de algún modo los sucesivos cambios “ascendentes” acaban dando lugar a un ciclo cerrado. Es decir, a pesar de la sensación de alejarse siempre más del origen, uno termina, para su propio sobresalto, exactamente donde había arrancado. En resumen, un bucle extraño es un bucle de realimentación paradójico con pasos de nivel (*ibid.*, pp. 101-2).

Ahora bien, ¿cuál es la relación de los bucles extraños con la conciencia? Para responder a esta pregunta, debemos conocer cómo concibe Hofstadter la relación entre el sistema nervioso y la mente en el ser humano. El autor compara la mente y el cerebro con “el auto-modificable software y el intacto hardware” (Hofstadter 1979, p. 686), respectivamente:<sup>9</sup> si bien somos capaces de realizar cambios en las reglas que operan a nivel mental, nuestras neuronas, en cambio, emplean siempre las mismas reglas para su funcionamiento. Mientras que nuestros pensamientos están a nuestro alcance, nuestras neuronas no lo están. Hofstadter clarifica esta distinción mediante otra ilustración de Escher: *Drawing hands* (figura 5.6). En esta obra, la mano derecha es la autora del dibujo que lleva a su existencia a la mano izquierda, pero, al mismo tiempo,

---

<sup>8</sup> Con el mismo fin, Hofstadter recurre también al “Canon per Tonos” de la Ofrenda Musical de J. S. Bach y al teorema de incompletitud de K. Gödel. No obstante, nos centraremos en los ejemplos proporcionados por las obras de Escher, pues consideramos que su lenguaje visual resulta más fácilmente asimilable para su empleo como ejemplos.

<sup>9</sup> Esta doble analogía posee un marcado tinte putnamiano.

la mano izquierda está realizando el dibujo de la mano derecha. La relación entre ambas constituye una “jerarquía enredada”, término que utiliza el autor para referirse a una jerarquía que alberga un bucle extraño (*e.g.*, Hofstadter 1979, pp. 10 y 689). Sin embargo, es Escher, como creador de la ilustración, quien provoca que el bucle extraño formado por las dos manos sea posible. Escher se encuentra en un nivel, intacto e invisible, que actúa como sostén para otro, auto-modificable y visible, en el que se halla la jerarquía enredada que alberga el bucle extraño. Hofstadter efectúa una analogía entre esta relación internivel y el vínculo entre los niveles neural y mental. Los numerosísimos símbolos presentes en el pensamiento conforman una maraña que se encuentra aislada respecto al nivel intacto que subyace a ella, lo que nos impide apreciar que dicho nivel existe. En él, encontramos una enorme cantidad de neuronas y axones relacionándose, también, de manera enredada. Así, “un enredamiento de software, el de los símbolos, está soportado por un enredamiento de hardware, el de las neuronas” (Hofstadter 1979, p. 691). Sin embargo, este último no transgrede la diferenciación entre niveles de carácter jerárquico previamente concebida, razón por la cual no constituye una *jerarquía enredada*. Sí lo hace, en cambio, el enredamiento que se produce entre los símbolos, y de ahí su aislamiento respecto al nivel inferior. Sobre los fundamentos de dicho enredamiento hablaremos a continuación.

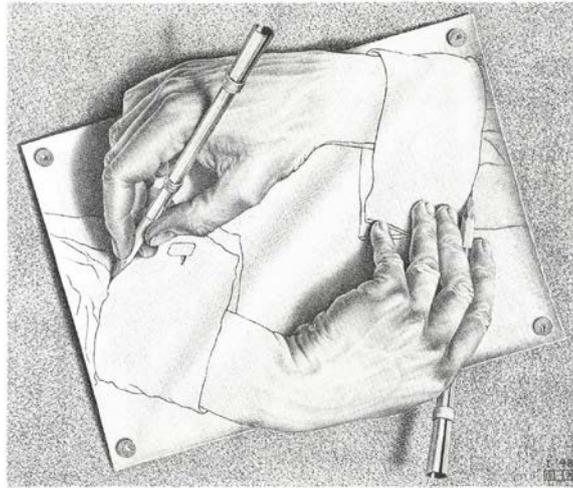


Figura 5.6. *Drawing hands*, de M. C. Escher. Tomado de Escher (1992).

Los organismos gozan de la habilidad consistente en la percepción y categorización de aquello que ocurre en su ambiente próximo. Esta habilidad, afirma Hofstadter, implica asimismo la percepción de algunos elementos internos. Una realimentación de vídeo, en cambio, puede recibir señales pero no puede percibir. En la percepción, *la entrada de señales viene seguida por una activación selectiva de símbolos que se hallaban en estado latente*. ¿Qué es un símbolo? Para nuestro autor, “debería ser concebido como una estructura física activable que constituye la forma en que el cerebro implementa una particular *categoría* o *concepto*” (Hofstadter 2007, p. 75). Y añade:

La idea que quiero expresar mediante la frase “un símbolo en el cerebro” es que cierta estructura específica dentro de tu cráneo [...] resulta activada cada vez que piensas en, pongamos, la Torre Eiffel. Esa estructura cerebral, cualquiera que fuere, es lo que yo llamaría tu “símbolo de la Torre Eiffel”.

También tienes un símbolo “Albert Einstein”, un símbolo “Antártida” y un símbolo “pingüino”, siendo el último algún tipo de estructura dentro de tu cerebro que resulta disparada cuando percibes uno o más pingüinos, o incluso cuando estás *simplemente pensando* [cursiva añadida] sobre los pingüinos sin percibir ninguno. También hay, en tu cerebro, símbolos para conceptos de acción como “kick”, “kiss” y “kill”, para

conceptos relacionales como “before”, “behind” y “between”, etcétera.<sup>10</sup> En este libro, entonces, los símbolos en un cerebro son las entidades neurológicas que corresponden a los conceptos, de la misma manera en que los genes son las entidades químicas que corresponden a los rasgos hereditarios (*ibid.*, p. 76).

Las numerosísimas señales que entran en un principio sufren un procesamiento que provoca la activación selectiva de otras. Estas son, a su vez, procesadas, lo cual conduce a otra activación selectiva de señales. Este mecanismo sigue operando mediante “una especie de proceso de canalización” (*id.*) que va estrechando las trayectorias cerebrales hasta activar, finalmente, un grupo de símbolos reducido. No obstante, el proceso incluye la actuación de otras señales, de origen interno, liberadas por causa de algunos símbolos que se asocian a las expectativas generadas en el individuo a raíz de acontecimientos vividos en el pasado. Estas señales interactúan con las que se transmiten hacia el interior, y de este contacto entre señales con sentidos opuestos surge la activación, ya mencionada, de unos pocos símbolos. Ahora bien, la percepción, no exclusiva de los seres humanos, viene acompañada en nuestra especie por dos cualidades. Una es la capacidad para establecer, mediante anidamiento, una jerarquía ilimitada de conceptos. La otra consiste en una memoria episódica más completa que la que pudiera tener otro animal, pues está dotada con un alto grado de precisión y detalle. Ambas cualidades permiten que, al percibirnos a nosotros mismos, surja el símbolo del “yo”: el más relevante y complejo que hay en un cerebro humano. Sin embargo, dicho símbolo no reside en una ubicación determinada del cerebro, ya que consiste, realmente, en un subsistema de símbolos que se despliega por todo el órgano. Este subsistema cerebral, sostiene Hofstadter, *constituye un nivel jerárquico propio y controla la activación individual de los símbolos que lo conforman.*

El símbolo del “yo” se desarrolla de manera progresiva a lo largo del tiempo. Esto sucede mediante observación y asimilación de las consecuencias

---

<sup>10</sup> Nota de traducción: los seis términos entrecomillados en esta oración han sido respetados en su forma inglesa por considerar que existe una manifiesta intención, por parte del autor, de mostrar palabras con fonética u ortografía similares dentro de cada trío de ejemplos.

de las acciones que son producto de su elección. El proceso tiene lugar una y otra vez. Dice Hofstadter:

A través de interminables exploraciones aleatorias [...], mi auto-símbolo adquiere lentamente una percepción concisa y valiosa hacia su naturaleza como elector y emisor de acciones, integrado en un mundo vasto, variopinto y parcialmente predecible.

Para ser más concreto: lanzo una pelota de baloncesto hacia una canasta, y gracias a hordas de eventos microscópicos en mis brazos, mis dedos, la revolución de la pelota, el aire, el aro, etcétera, de todos los cuales soy inconsciente, o fallo o acierto mi lanzamiento de gancho. Este diminuto tanteo del mundo, repetido cientos o miles de veces, me informa cada vez más fielmente sobre mi nivel de habilidad como jugador de baloncesto (y también me ayuda a decidir si me gusta ese deporte o no) (*ibid.*, pp. 183-4).

Y añade:

De forma similar, mis actos sociales inducen reacciones por parte de otros seres sensibles. Esas reacciones rebotan hacia mí y las percibo en términos de mi repertorio de símbolos, y de esta manera me percibo indirectamente a mí mismo a través de mi efecto sobre otros. Estoy desarrollando mi sentido de quién soy a ojos de los demás. Mi auto-símbolo se está formando a partir de un vacío inicial (*ibid.*, p. 184).

Así, el “yo” evoluciona poco a poco en virtud de un bucle en el que las acciones ocasionadas por símbolos tienen consecuencias que activan símbolos. Además, la evolución trae consigo progresivamente la convergencia, la estabilización y la fijación de ese “yo”.

En definitiva, el cerebro humano, según Hofstadter, al percibir la actividad de sus propios símbolos, efectúa un *giro jerárquico*. Esto no ocurre en una realimentación de vídeo, donde no hay símbolos ni percepción. Sin embargo, nuestro cerebro *no puede observar la actividad inferior*, que tiene lugar a nivel neuronal. Ambos aspectos denotan que nos hallamos ante un *bucle extraño*. Nuestro “yo”, concluye el autor, se desarrolla en el cerebro a partir de semejante patrón.

La visión de Hofstadter resultará esencial para el planteamiento de nuestra tesis  $\delta'$ . Sin embargo, no coincidimos con algunas de las posiciones que

defiende en los dos libros anteriormente mencionados, en especial con dos de ellas. Por un lado, no estamos de acuerdo con él cuando sostiene que nuestro “yo” es “un mero epifenómeno pasivo que surge de niveles inferiores” (*ibid.*, p. 205), “una trampa en la que todos los humanos caemos” pero que “es lo que nos hace humanos” (*id.*), “un *mito* auto-reforzador” (*ibid.*, p. 291), “una ilusión tremendamente eficaz” (*ibid.*, pp. 291-2), o, incluso, “una alucinación percibida por una alucinación” (*ibid.*, p. 293). En oposición a estas afirmaciones, propondremos (en la siguiente sección) que el bucle extraño que da lugar a ese “yo” goza de poder causal precisamente sobre esos niveles inferiores a los que alude Hofstadter, razón por la cual no podría constituir un epifenómeno (por definición) ni, desde luego, una trampa, una ilusión, etc.

También diferimos de Hofstadter en cuanto a su tratamiento del libre albedrío, de corte determinista. Inicialmente, en *Gödel, Escher, Bach* sostiene que el libre albedrío se sustenta en un bucle extraño. También que “es un resultado de la interacción entre el auto-símbolo (o subsistema) y los otros símbolos del cerebro” (Hofstadter 1979, p. 710) y que lo sentimos a raíz de un balance en el que participan tanto la ignorancia como el conocimiento de uno mismo. Sin embargo, en *I Am a Strange Loop* nuestro autor se muestra más categórico y extrema sus expresiones. Así, tilda al mencionado problema de “vaca sagrada” y afirma que le horroriza (Hofstadter 2007, p. 339). Además, con notable desdén, asegura que se propone “despacharlo [...] lo más rápidamente posible” (*id.*). No en vano, dedica al libre albedrío solamente tres páginas (*ibid.*, pp. 339-41), en las cuales carga contra su existencia. Su línea de razonamiento queda patente en los siguientes fragmentos:

[P]odría escoger no tomar una segunda ración de fideos aunque a mí – o, más bien, a una parte de mí – aún me apeteciera un poco, porque hay *otra* parte de mí que no quiere que gane peso, y ocurre que (esta noche) la parte que vigila el peso tiene más votos de los que tiene la parte glotona. Si no los tuviera, entonces perdería y mi glotón interno vencería, y eso estaría bien – pero, en

cualquier caso, mi *no libre albedrío* [cursiva añadida] saldría victorioso y yo seguiría el deseo dominante en mi cerebro.<sup>11</sup>

Sí, ciertamente, tomaré una decisión, y lo haré llevando a cabo una especie de votación interna. El recuento de los votos dará un resultado, y [...] un lado saldrá vencedor. ¿Pero dónde hay “libertad”<sup>12</sup> en todo esto? (*ibid.*, p. 340).

[También hay] factores externos que actúan como constricciones o, dicho más metafóricamente, que desempeñan el papel de setos en el vasto laberinto de la vida en el que estamos atrapados. Gran parte de la vida es increíblemente aleatoria, y no tenemos control sobre ella. Podemos anhelar todo lo que queramos, pero gran parte del tiempo nuestra voluntad es frustrada (*ibid.*, p. 341).

Obsérvese que los argumentos de Hofstadter recuerdan a los que emplean los compatibilistas clásicos: el agente goza de la capacidad para hacer lo que desea, cosa que logrará si no hay nada que se lo impida. Sin embargo, mientras que, para el compatibilismo clásico, ser libre consiste precisamente en lo que acabamos de decir, nuestro autor piensa todo lo contrario en tanto en cuanto niega que exista el libre albedrío. Por eso, su posición parece más cercana, en realidad, a la del determinismo duro, según el cual el libre albedrío no existe porque, en primer lugar, no es compatible con el determinismo, y, en segundo lugar, el determinismo es cierto.<sup>13</sup>

A pesar del rechazo de Hofstadter hacia el libre albedrío, la postura que defiende acerca de la relación entre el cerebro y la mente por medio de un bucle extraño ha sido empleada, paradójicamente, para defender su existencia. Así ocurre con Michael Morden (1990), si bien es el único caso del que

---

<sup>11</sup> El Barón de Holbach ya hablaba, en el siglo XVIII, de pasiones dominantes que imperan sobre las restantes.

<sup>12</sup> Nota de traducción: lo que aquí se ha traducido como “libertad” es expresado por Hofstadter mediante el vocablo “freeness” (el autor también emplea las comillas).

<sup>13</sup> Acerca del compatibilismo clásico y del determinismo duro, véanse, respectivamente, las secciones 1.2. y 1.3.

tenemos constancia.<sup>14</sup> <sup>15</sup> Morden ofrece un planteamiento compatibilista que gira en torno a la fórmula: “Soy (suficientemente) libre cuando soy una de las causas de mi propia conducta” (*ibid.*, p. 72). Los seres humanos, sostiene, tienen exclusividad en lo referente a la capacidad para dirigir la conciencia sobre sí misma. No en vano, “pueden darse cuenta de que tienen ciertas creencias, deseos, y [...] son capaces de comprender las causas de sus acciones” (*ibid.*, p. 61). Dicha capacidad les permite gozar de libre albedrío. Morden afirma que este es concebido en *Gödel, Escher, Bach* como resultado de la interacción que se produce entre el concepto que una persona tiene de sí misma y las causas que subyacen a sus actos, y que dicha concepción de Hofstadter apoya su propia idea. Ahora bien, Morden resalta que nuestro “yo” constituye una condición necesaria pero no suficiente para nuestro comportamiento, ya que existen otras influencias necesarias. Para que nuestra conducta pueda ser considerada como libre, nuestra comprensión de las influencias que la afectan debe prevalecer sobre el resto de condiciones necesarias, pues de lo contrario no cabe afirmar que dicha comprensión constituya una condición necesaria.<sup>16</sup> También resulta imprescindible que la cadena causal no rodee al agente (*e.g.*, si otra persona levanta el brazo del agente),<sup>17</sup> sino que lo atraviese. Si el agente actúa conforme a razones, este requisito se satisface. No necesitamos el indeterminismo, sostiene el autor; lo

---

<sup>14</sup> Posteriormente, este artículo ha sido estudiado a fondo por Bertrán (1993), pero se trata de un trabajo de revisión y no se presentan nuevas aportaciones.

<sup>15</sup> Obviamente, el autor sólo tiene en cuenta las ideas de Hofstadter defendidas en *Gödel, Escher, Bach*. Nótese que *I Am a Strange Loop* es de publicación bastante posterior (2007) al artículo aquí referenciado.

<sup>16</sup> Un detractor de esta afirmación podría quizá argumentar que si la comprensión de las influencias que afectan a nuestra conducta prevalece sobre el resto de condiciones necesarias, estas dejan de ser, precisamente, necesarias. Sin embargo, si no existieren tales condiciones, no podría darse una comprensión de ellas, lo cual las convierte en necesarias de todos modos.

<sup>17</sup> Morden toma este ejemplo (y así lo hace saber) del capítulo 4 de *Analytical Philosophy of Action*, de Arthur Danto (1973, pp. 79-115).

que necesitamos es “sólo una cadena causal que regrese sobre sí misma por medio de la precisa autoconsciencia que es crucial para nuestra naturaleza como personas” (*id.*).

Aunque, como ya hemos dicho, Morden se sirve de los bucles extraños para defender el compatibilismo, por nuestra parte pensamos que cabe una salida incompatibilista libertarista en virtud de una modalidad de causalidad mental que denominaremos *poda deliberativa*.

## 5.6. La “poda deliberativa” como modalidad de causalidad mental

Nuestra propuesta se fundamenta en la influencia causal de un bucle extraño sobre el cerebro humano, concebido este último como sistema no integrable. Dicho bucle constituiría un patrón abstracto presente en procesos de deliberación, mientras que la influencia causal sobre el sistema nervioso procedería por medio de la poda sináptica. Como ya apuntábamos al principio del capítulo, nos limitaremos a plantear una *posibilidad inteligible* de conciliación entre los postulados materialistas y el libertarismo. Los aspectos fundamentales de nuestra aproximación son los siguientes:<sup>18</sup>

[I] *Deliberación iterativa*. Un individuo humano está sujeto durante toda su vida (aunque con mayor intensidad en su infancia temprana) a procesos de *poda semántica* (tesis  $\delta$ ), en virtud de los cuales adquiere una serie de habilidades supeditadas a estándares semánticos por medio de una poda sináptica en las rutas neurales y neuromusculares cuyos *outputs* se adecuan a dichos estándares (véase sección 5.2.). En concreto, la fijación de rutas correlacionada con la asimilación del lenguaje ordinario dispone al individuo al uso de estas en

---

<sup>18</sup> Como ya sucedía con la tesis  $\delta$ , emplearemos principalmente el presente de indicativo en aras de una mayor comodidad de lectura. Rogamos al lector que, en todo momento, recuerde que se trata de una aproximación hipotética.

procesos de deliberación (cf. Moya 2011, p. 199), recurrentes a partir de la adolescencia. En la deliberación práctica, en concreto, el individuo analiza con cuidado y detalle los deseos, creencias e intenciones que intervienen en relación a una determinada decisión vinculada a una acción. Cuando la interacción entre estos estados intencionales no resulte excesivamente compleja, la decisión podrá alcanzarse tras un solo ciclo deliberativo, tras lo cual el individuo ejecutará esta por medio de una acción.<sup>19</sup> En cambio, si la interacción es compleja, la deliberación se convierte en un mecanismo que opera por medio de *i* iteraciones (donde *i* corresponde al número de ciclos deliberativos efectuados con antelación a la decisión) y que conduce a decisiones, es decir, decisiones a medio o largo plazo, es decir, a *decisiones distales*. Este tipo de deliberación práctica, que podríamos denominar “deliberación práctica iterativa” (de ahora en adelante, simplemente *deliberación iterativa*), es el que consideraremos en lo sucesivo.

[II] *El cerebro como sistema no integrable*. Tal y como Bishop sostiene (véase sección 5.4.), el cerebro presenta los cinco atributos que caracterizan a los sistemas no integrables: propiedades de carácter emergente, numerosas partículas, actuación colectiva, estructura y orden considerables, y evolución irreversible. Por ello, supondremos que este órgano constituye un sistema no integrable englobado en un sistema mayor y de carácter complejo: el ser humano en su conjunto. Como cualquier otro sistema no integrable, su descripción más adecuada se realizará por medio de distribuciones que no son reducibles al comportamiento individual de los elementos que las componen, dado que las interacciones resultan imprescindibles para la mencionada descripción. Considerando que las interacciones fundamentales para el funcionamiento específico del cerebro son las sinapsis, la mejor descripción para estudiar este órgano será probablemente aquella que tenga a las neuronas como elementos individuales (no a las partículas) y a la distribución de los

---

<sup>19</sup> Podría ser también una cadena de acciones. No obstante, en búsqueda de una mayor claridad, concebiremos una única acción.

botones sinápticos <sup>20</sup> de estas (de ahora en adelante, *distribución sináptica*) como modelo fundamental de descripción.

[III] *Correlación entre lo neural y lo semántico*. Dentro del cerebro existen estructuras correlacionadas con categorías semánticas o conceptos (se trata de los “símbolos” de Hofstadter). Esta afirmación parece verse respaldada por el trabajo de Quiroga *et al.* (2005), donde se demuestra que en el lóbulo medial temporal humano tiene lugar una activación neuronal selectiva ante la presencia de señales muy distintas pero relativas al mismo concepto. Por ejemplo, los autores hallaron una misma respuesta ante fotografías de Halle Berry que diferían en varios aspectos: ángulo de la toma, con o sin gafas de sol, peinado, etc. También ante un dibujo de la actriz, ante fotografías donde aparecía disfrazada de Catwoman e, incluso, ante su nombre escrito. A juicio de los autores, sus “datos son compatibles con una representación abstracta de la identidad del individuo u objeto mostrado”, que “podría ser importante en el almacenamiento de recuerdos a largo plazo” (*ibid.*, p. 1106).

[IV] *Jerarquización de categorías semánticas*. Las estructuras neurales correlacionadas con categorías semánticas interaccionan para constituir vínculos jerárquicos. El modo en que esto ocurre en el cerebro se asemeja en su esencia al establecimiento de correlaciones espaciales en otro tipo de sistemas no integrables: los gases moleculares densos. El siguiente fragmento describe cómo las colisiones entre partículas conducen a la formación de correlaciones espaciales en estos gases:

Las colisiones son frecuentes en gases densos y las correlaciones espaciales provocadas por las colisiones acoplan cada partícula con muchas otras partículas (posiblemente todas) del gas. Es este acoplamiento debido a las correlaciones el que conduce al comportamiento colectivo responsable de que las partículas de gas se reúnan en estructuras coherentes en lugar de

---

<sup>20</sup> Los botones sinápticos son las estructuras terminales entre las cuales se produce la transmisión de señales en la sinapsis. Hay botones tanto axónicos (presinápticos) como dendríticos (postsinápticos).

extenderse uniformemente por todo el volumen. Serían ejemplos la turbulencia y las ondas de choque.

Para ver cómo se desarrollan estas correlaciones, empiece con las partículas del gas antes de que hayan interactuado unas con otras. Según comienzan colisionando, las primeras interacciones establecen correlaciones binarias entre partículas. A medida que las interacciones persisten, empiezan a aparecer correlaciones ternarias. El proceso continuará estableciendo combinaciones cuaternarias, y así sucesivamente a través de correlaciones  $N$ -arias [...]. La progresión desde las correlaciones de orden inferior (que aparecen primero) a las correlaciones de orden superior (que aparecen después) corresponde a un ordenamiento temporal natural de la evolución de los estados del gas. Las correlaciones y otros efectos colectivos pueden rivalizar con, o superar, el papel de las trayectorias individuales de las partículas (Bishop 2004, pp. 20-21).

De manera análoga, las neuronas del cerebro interactúan por medio de sinapsis y, en virtud de procesos de poda semántica, van formando correlaciones persistentes en las que unas con otras se acoplan, mediante impulsos nerviosos recurrentes, para formar estructuras coherentes desde el punto de vista funcional. Las correlaciones de órdenes más bajos conducen a estructuras neurales relativamente simples, asociadas con categorías semánticas sencillas. Algunas de estas estructuras conforman otras más complejas, asociadas a categorías de nivel jerárquico más alto (por ejemplo, frases sencillas y conceptos abstractos) por medio de correlaciones de órdenes superiores. Correlaciones de orden creciente continúan permitiendo el anidamiento de ciertas categorías hasta llegar a un punto en el que componen una estructura coherente asociada con nuestro “yo” o identidad personal. El funcionamiento de esta supraestructura corresponde a un comportamiento neural colectivo y a gran escala en el cerebro.

[V] *Bucle extraño del “yo”*. La identidad personal o “yo” constituye un supranivel jerárquico de carácter semántico, y consiste en el concepto que sobre sí mismo alberga el agente. Según sostiene Hofstadter, dicho autoconcepto ha venido desarrollándose ininterrumpidamente, desde el principio de la vida del individuo, en virtud de un bucle extraño consistente en: 1) la percepción del agente sobre sí mismo a partir de las consecuencias de sus acciones, y 2) la realización de nuevas acciones a partir de dicha percepción.

[VI] *Percepción semántica.* Para que la deliberación iterativa comience, es necesaria una percepción de señales que guarden contenido semántico. Las señales penetran, a través del sistema sensorial, hacia las estructuras cerebrales responsables de la sensación y la percepción (por ejemplo, los córtex visual y auditivo primarios), y activan las transmisiones sinápticas en aquellas rutas neurales que la poda semántica ya correlacionó con ellas en situaciones de aprendizaje. Estas vías neurales selectivamente activadas corresponden a categorías semánticas amplias, pero envían nuevas señales, esta vez internas (mediante impulsos eléctricos), provocando que, de entre todas las vías correspondientes a categorías más específicas, se seleccionen las que se mejor se adecuan a las señales captadas sensorialmente en un principio. Este mecanismo descendente se repite varias veces, e interactúa con otro de carácter ascendente en el que se transmiten señales internas asociadas a expectativas surgidas a raíz de episodios anteriores. El resultado es que una única ruta neural (o quizá unas pocas), muy específicamente correlacionada con las señales iniciales, es la que acaba por efectuar transmisiones sinápticas. Si tomamos prestado por un momento el ejemplo de Halle Berry y lo aplicamos al mecanismo de percepción que acabamos de describir, una secuencia jerárquicamente descendente de categorías conceptuales podría ser la siguiente: persona – mujer – actriz – actriz de la película *Catwoman* – Halle Berry.<sup>21</sup>

[VII] *Conflicto intencional.* Las categorías conceptuales correspondientes a las rutas neurales activadas en la fase perceptiva pueden interactuar con deseos, creencias y/o intenciones del agente que tienen algún tipo de relación semántica con ellas (por ejemplo, ver una fotografía de Halle Berry disfrazada de Catwoman puede interactuar con una intención de abrir una tienda de disfraces como negocio familiar). De esta situación puede surgir un conflicto entre estados intencionales contrapuestos (e.g.: “pretendo abrir una tienda de

---

<sup>21</sup> Por supuesto, la longitud y el contenido de la secuencia son arbitrarios. Podría ser más corta o (probablemente) más larga y estar compuesta por conceptos distintos a alguno de los que en ella aparecen.

disfraces y deseo hacerme rico, *pero* creo que mi familia está en contra de abrir negocios familiares y deseo llevarme bien con mi familia”). Ante dicho conflicto debe tomarse una decisión. Entonces dará comienzo un proceso de deliberación, en el que el individuo ya no percibirá señales externas sino internas, muchas de ellas relativas a los diferentes estados en conflicto. Estos deseos, creencias e intenciones surgen del anidamiento jerárquico de categorías semánticas sencillas (referentes a objetos, cualidades, acciones, relaciones, etc.) que, a nivel neural, corresponden a vías relativamente simples y ya fijadas mediante poda semántica en procesos de aprendizaje. El anidamiento semántico, por su parte, corresponde a correlaciones de orden creciente que dan lugar a redes neurales complejas. Dicho anidamiento, sin embargo, no equivale a un anidamiento estructural, de manera que estas redes se encuentran entrelazadas intrincadamente por todo el cerebro, conformando una superred neural. Además, cada neurona puede participar en más de una red neural, es decir, no se ve comprometida necesariamente por una única categoría semántica.

[VIII] *Iteración 1.* Sea  $d_0$  la distribución sináptica que presenta el cerebro en el momento en que se inicia la primera iteración deliberativa. Los estados intencionales en conflicto interaccionan en diversas combinaciones que conducen a un cierto número de *posibilidades alternativas* de acción ( $P_1, P_2, \dots, P_n$ ). Consideraremos que estas  $P$  son equiprobables en cuanto a sus posibilidades *a priori* de ser finalmente seleccionadas. Al deliberar, el individuo recrea internamente las posibles consecuencias de cada  $P$ , para lo que se sirve de analogías formadas a partir de recuerdos episódicos con los que el conflicto presente guarda alguna similitud. Dichos recuerdos, vividos por el individuo en primera persona, traen asociados emociones y sentimientos que el individuo experimentó cuando sucedieron los episodios, de manera que la recreación de consecuencias análogas genera emociones y sentimientos igualmente análogos. Estas señales internas, que constituyen lo que Damasio denomina “marcadores somáticos” (véase apartado 4.1.3.), pueden ser gratas o desagradables, llevando,

respectivamente, a aceptar ciertas  $P$  o a rechazarlas. Los marcadores somáticos actúan con celeridad, aumentando la simplicidad y la eficacia del proceso deliberativo. Las  $P$  que pasan el filtro son entonces evaluadas según la *adecuación de sus posibles consecuencias* al concepto que de sí mismo tiene el agente.<sup>22</sup> Este autoconcepto o “yo” establece un estándar normativo que actúa como constrictión contextual de carácter interno, de manera que las  $P$  que se adecuan a sus reglas de significado son seleccionadas en detrimento del resto.<sup>23</sup> Cuando el número de  $P$  seleccionadas es mayor que 1, la primera iteración no ha sido suficiente para desencadenar una determinada  $A$ , siendo necesario un nuevo ciclo deliberativo, o iteración 2. No obstante, este primer ciclo marca el inicio de un mecanismo de poda sináptica en el que se refuerzan las sinapsis de las redes neurales correspondientes a los estados intencionales cuya interacción conduce a las  $P$  seleccionadas. El refuerzo provoca que estas  $P$  aumenten su probabilidad de selección. Por contra, las sinapsis relacionadas con las  $P$  descartadas se debilitan y dichas posibilidades ven disminuida su probabilidad. Se rompe, así, la equiprobabilidad inicial (figura 5.7). La poda supone, además, un efecto colectivo: la redistribución de botones sinápticos, que, finalizada la iteración 1, da como resultado una distribución sináptica transitoria.

---

<sup>22</sup> Nótese la similitud con la relación jerárquica que Harry Frankfurt (1971) establece entre los “deseos de primer orden” y los “deseos de segundo orden” (véase sección 1.2). Los deseos de segundo orden, que surgen de una autoevaluación reflexiva, controlan a los de primer orden, y estos serán libres cuando sean coherentes con los de segundo orden.

<sup>23</sup> Como el lector habrá observado, esta explicación interrelaciona fuertemente la cognición, la emoción y la memoria. Por esta razón, pensamos que es compatible con el concepto de metaestabilidad, del cual hablábamos en el apartado 3.3.5.



*sináptica*, provoca que se formen correlaciones nuevas de carácter impredecible y que la distribución evolucione por medio de efectos colectivos aleatorios. Esto podría conducir al surgimiento de nuevas  $P$  por la aparición de nuevos estados intencionales y/o la recombinación de algunos de los ya existentes, y, en consecuencia, a la modificación de las probabilidades correspondientes a cada  $P$  presente al comienzo del interciclo (figura 5.7.). Hay una cierta analogía entre la deriva sináptica y la deriva genética, en la medida en que esta última, que constituye uno de los cuatro factores intervinientes en la evolución biológica (junto a la selección natural, la mutación y la migración), provoca cambios aleatorios en las frecuencias alélicas. Sin embargo, la deriva genética disminuye la variabilidad genética en una población biológica, mientras que la deriva sináptica, tal y como la acabamos de plantear, es capaz de incrementar la variabilidad en la “población” de  $P$ . Por otro lado, si el interciclo es lo suficientemente prolongado, el efecto estocástico de la deriva sináptica puede verse amplificado por la intervención de otros mecanismos, como la poda semántica (en nuevos aprendizajes), la poda deliberativa (en deliberaciones iterativas distintas a la que está en curso) o la neurogénesis. De cualquier modo, una vez finalizado el interciclo 1, tenemos una nueva distribución sináptica transitoria.

[X] *Iteración 2.* En la segunda iteración, el individuo recrea nuevas consecuencias posibles derivadas de cada  $P$  superviviente, si bien se trata, en general, de consecuencias a medio y largo plazo, derivadas de situaciones hipotéticas más complejas. En el caso  $b$ , el agente puede haber registrado nuevos recuerdos episódicos, asociados a sus respectivos marcadores somáticos. De manera similar al proceso seguido en la iteración 1, los nuevos marcadores somáticos actúan como nuevo filtro para las  $P$  aún activas y aumentan, así, la eficacia y simplicidad de la deliberación. De cualquier forma, tanto en el caso  $a$  como en el  $b$  las  $P$  supervivientes son sometidas a la evaluación normativa de sus posibles consecuencias conforme al estándar semántico establecido internamente por el autoconcepto o “yo” del individuo.

Dicho autoconcepto, por cierto, puede haber sufrido alguna modificación, si el interciclo 1 ha sido lo suficientemente largo, debido a la percepción de las consecuencias de nuevas acciones. En cualquier caso, para las  $P$  que resultan seleccionadas, tiene lugar un refuerzo de las sinapsis de las redes neurales correspondientes a los estados intencionales cuya interacción conduce a dichas posibilidades, lo cual provoca que estas aumenten sus probabilidades de selección en detrimento del resto. Para las  $P$  descartadas, en cambio, tiene lugar un debilitamiento de las sinapsis. Esta poda sináptica supone un efecto colectivo: la redistribución de botones sinápticos que, finalizada la iteración 2, conduce a una nueva distribución transitoria.

[XI] *Interciclo 2.* Al igual que en el primer interciclo, las constricciones normativas internas establecidas por el “yo” dejan de actuar y la distribución sináptica evoluciona aleatoriamente, con efectos colectivos impredecibles. Las probabilidades correspondientes a cada  $P$  pueden verse modificadas y también pueden surgir nuevas  $P$ . Finalizado el interciclo 2, el cerebro muestra una nueva distribución transitoria.

[XII] *Resolución del conflicto.* En caso de ser necesarias más iteraciones, en cada una de ellas, y tras un interciclo, el proceso se repite: el agente recrea nuevas consecuencias, derivadas de situaciones más complejas que las recreadas en la iteración precedente, pero los marcadores somáticos y la adecuación de dichas consecuencias al estándar normativo establecido por el “yo” establecen constricciones que simplifican la operación. Al reforzar progresivamente algunas redes neurales y debilitar otras, la poda sináptica va aumentando las probabilidades de selección de ciertas  $P$ , mientras que las probabilidades de las demás van disminuyendo con las sucesivas iteraciones hasta ser iguales a cero. Así, una vez finalizada cierta iteración  $i$ , resta una única  $P$ , con una probabilidad de selección igual a 1 (figura 5.7.). Esta se convierte, así, en la *decisión* (distal) al conflicto que desencadenó la deliberación, y, llegado el momento, conducirá a la acción vinculada a ella. Por su parte, los deseos, creencias e intenciones (estados intencionales) cuya interacción conduce a la

posibilidad vencedora (la decisión) constituyen las *razones de la acción*. En cuanto a la poda sináptica, supone, como efecto colectivo, la redistribución de botones sinápticos que, una vez finalizada la iteración  $i$ , conduce a  $d_i$  (distribución sináptica final).

[XIII] *Poda deliberativa*. En definitiva, a lo largo de  $i$  iteraciones se reafirma progresivamente la adecuación al estándar normativo interno (establecido por el autoconcepto) de las consecuencias posibles recreadas por el agente en relación a la  $P$  ganadora (la decisión ulterior). A escala neural, esta reafirmación está correlacionada con una poda sináptica que actúa de forma paulatina y sostenida, consiguiendo un refuerzo cada vez mayor de las redes neurales correspondientes a los estados intencionales vencedores del conflicto. Todo este proceso, al que denominaremos *poda deliberativa*, implica un doble vínculo de emergencia contextual. En primer lugar, la descripción del comportamiento individual de las neuronas de un cerebro humano, que es un sistema no integrable, resulta necesaria pero insuficiente para explicar la evolución de la distribución sináptica, para lo cual debe acudir al estándar semántico establecido por el contexto interno. En segundo lugar, la descripción de la distribución sináptica es necesaria pero insuficiente para explicar el proceso de deliberación, para lo cual debe acudir nuevamente al estándar semántico mencionado. En ambos casos, debe acudir también al componente fenomenológico presente en las emociones que participan a través de los marcadores somáticos.

[XIV] *Flecha del tiempo y entropía*. Desde el punto de vista termodinámico,  $d_0$  describe una fase del sistema, es decir, una estructura macroscópica con propiedades físicas y químicas que le son particulares. El surgimiento del conflicto intencional constituye una modificación contextual que desencadena una transición de fase, de forma que  $d_0$  comienza a evolucionar, desde la iteración 1, por medio de cambios neuronales inducidos por la poda sináptica. Cuando termina la iteración  $i$ , la poda finaliza y se llega a una nueva fase termodinámica, a la que corresponde una distribución  $d_i$ . La

descripción de la transición de fase que conduce desde  $d_0$  hasta  $d_i$  sólo es realizable en términos estadísticos. Dicho proceso no es reducible al comportamiento individual de las neuronas, y sólo es comprensible si se tienen en cuenta las correlaciones (interacciones) persistentes que se establecen entre ellas. Además, en la transición de  $d_0$  a  $d_i$  se produce una rotura de simetría: al igual que “observamos que un vaso que cae se rompe pero jamás observamos que los añicos se recompongan reconstruyendo el vaso original” (Cuesta 2006, p. 13), tampoco es posible reconstruir  $d_0$  a partir del resultado final  $d_i$ . Aparece, así, una flecha del tiempo termodinámica, y esta conlleva un aumento de entropía entre las dos fases. Dado que la entropía es la cantidad de energía que no está disponible para realizar trabajo, y que una fuerza realiza trabajo cuando logra desplazar un cuerpo, podemos afirmar que un aumento de entropía en el sistema equivale a una disminución de la cantidad de energía disponible para efectuar desplazamientos internos. En el caso que nos ocupa, los intercambios de energía que se producen a nivel interno son de naturaleza eléctrica (a través del impulso nervioso) y química (por medio de los neurotransmisores), mientras que los desplazamientos más importantes son aquellos que sufren los botones sinápticos al redistribuirse durante la poda sináptica. Por tanto, podemos particularizar la afirmación anterior diciendo que el aumento de entropía entre  $d_0$  y  $d_i$ , asociado a la flecha del tiempo entre ambas distribuciones, equivale a una disminución de la energía eléctrica y química disponible para redistribuir los botones sinápticos durante la poda. De esta manera, la redistribución va menguando a medida que el proceso deliberativo avanza hacia su final.

[XV] *Entropía y orden*. Más arriba indicábamos que cada neurona puede participar en más de una red neural. El refuerzo efectuado en la poda deliberativa progresa de dos maneras complementarias. Las sinapsis que cada neurona dispone en las redes neurales que, dada su inadecuación al estándar semántico, son descartadas por constricción contextual interna, desaparecen paulatinamente con la retracción de ramales de su axón y/o la supresión de

espinas dendríticas de las células postsinápticas con las que está vinculada. Así, disminuye poco a poco la cantidad de axones que van a parar a una misma neurona postsináptica y el número de neuronas postsinápticas conectadas a un mismo axón. Además, se forman sinapsis adicionales entre las neuronas que pertenecen a las redes correspondientes a los estados intencionales ganadores en el conflicto deliberativo. En consecuencia, se forman circuitos con redundancia y solapamiento gradualmente menores, y adaptados a la perspectiva interna única del individuo. Ahora bien, la reducción de la redundancia y el solapamiento implica el aumento del orden del sistema, lo que parece contradecir una concepción corriente de la entropía según la cual un incremento positivo de esta equivale a la generación de desorden. Sin embargo, el aumento de entropía puede producir estructuras ordenadas, como bien explica Cuesta (*ibid.*). Así sucede, por ejemplo, con la formación de una red cristalina regular de esferas duras:

La explicación de este fenómeno es bien conocida por los fruteros [...]. Ellos saben de siempre que la mejor manera de empaquetar naranjas (o cualquier otra fruta más o menos esférica), es decir, la forma en que ocupan menos espacio, es apilándolas en una red ordenada. Lo que esto significa es que en una red ordenada, el volumen accesible medio que corresponde a cada partícula es mayor que en una fase desordenada (*ibid.*, pp. 14-15).

[E]l volumen accesible total depende de la configuración concreta de las esferas, y su promedio, por tanto, dependerá de la fase en la que se encuentren. Así pues, sí puede ocurrir (y de hecho ocurre) que en la fase cristalina las esferas dispongan de más volumen accesible que en la fase fluida (*ibid.*, p. 15).

[M]ás volumen accesible significa un mayor número de microestados, lo que a su vez significa [...] mayor entropía; la fase ordenada tiene más volumen accesible por partícula, luego *la fase ordenada tiene mayor entropía (id.)*.

Podemos trazar una analogía entre este caso y el aumento de entropía en las redes neurales. Debido a que la poda sináptica reduce la cantidad tanto de axones que desembocan en una misma neurona postsináptica como de neuronas conectadas con un mismo axón, nos encontramos con dos efectos

interesantes. En primer lugar, los botones sinápticos tienden a acumularse en pequeños grupos muy concentrados en lugar de repartirse de forma irregular en el espacio, que es lo que sucede en redes redundantes y solapadas. En segundo lugar, los axones trazan mayoritariamente vías unidireccionales (neurona – neurona) en detrimento de las ramificaciones multidireccionales, predominantes previamente a la poda. Ambos efectos implican un mayor orden sistémico en  $d_i$  del que había en  $d_0$ . Sin embargo, también implican un mayor volumen accesible medio por neurona (*i.e.*, el volumen del que cada una goza, en promedio, para efectuar movimientos) y, por tanto, una mayor entropía. En consecuencia, el aumento de entropía y el del orden pueden correlacionarse en el cerebro.

[XVI] *Indeterminismo.* Teniendo en cuenta las condiciones impuestas por la naturaleza en el momento de comenzar la deliberación iterativa, los acontecimientos futuros pueden desarrollarse de diversos modos. Este indeterminismo presenta tres dimensiones complementarias entre sí. En primer lugar, existe una incertidumbre del agente ante su propio proceder futuro, lo que le motiva para continuar deliberando (dimensión subjetiva). En segundo lugar, la distribución sináptica  $d_0$  y las leyes de la naturaleza, aun siendo necesarios para alcanzar la distribución sináptica  $d_i$ , no conducen inevitablemente a esta (dimensión física). Y en tercer lugar, la interacción entre factores normativos (en las iteraciones) y aleatorios (en los interciclos) hace que sea impredecible el número de iteraciones necesarias para cerrar el conflicto intencional (dimensión cíclica).<sup>25</sup>

[XVII] *Decisión distal y acción.* La reconfiguración neural progresivamente efectuada por la poda deliberativa marca una especie de impronta memorística que permite que, llegado el momento de ejecutar la decisión, la acción se lleve a cabo de manera eficaz y cuasi automática. Dicho de otro modo: a través de la poda sináptica, la decisión queda “fijada” en los

---

<sup>25</sup> En la sección 5.9. profundizaremos acerca del indeterminismo en la poda deliberativa. Considérese este párrafo como un mero avance.

tejidos biológicos en espera de su ejecución oportuna. Al percibir una determinada señal (o señales) que, por su contenido y su significado, actúe como estímulo desencadenante de la acción, la red neural correspondiente a la decisión (o sea, a la *P* vencedora) se comunicará con las estructuras cerebrales motoras adecuadas (el área motora suplementaria o el área de Broca, por ejemplo), y estas, a su vez, activarán las vías neuromusculares apropiadas para conectar con los músculos, órganos efectores de la acción. Una vez ejecutada esta, traerá consigo ciertas consecuencias. El agente percibirá estas, y al hacerlo actualizará su autoconcepto (corroborándolo o modificándolo).

[XVIII] *Balance entre neurogénesis y poda sináptica.* Como ya señalamos al hablar de la poda deliberativa, la cantidad global de sinapsis en el sistema nervioso varía a lo largo de la vida del agente al cambiar el balance entre la formación de nuevas conexiones sinápticas (en la neurogénesis) y la eliminación de otras ya existentes (en la poda sináptica). La poda deliberativa puede influir en este “balance sináptico” desde la adolescencia, etapa vital en la cual el individuo comienza a estar lo suficientemente preparado biológica y psicológicamente como para efectuar deliberaciones iterativas. También son dignas de consideración las diferencias biológicas y sociológicas entre individuos. Por ejemplo, hay profesionales que, debido al trabajo que desempeñan, se ven empujados a deliberar con mayor frecuencia y complejidad que otros (al menos en principio). Este sería el caso de jueces, políticos, filósofos, etc.

[XIX] *Tesis final.* En consecuencia con todo lo anterior, proponemos emprender una vía de investigación, complementaria a la que corresponde a la tesis  $\delta$ , en la que se establezca un nuevo vínculo concreto entre el proceso neurobiológico de eliminación (poda) de sinapsis y el concepto de causalidad mental establecido en la tesis  $\gamma$ . Recordemos esta última:

#### Tesis $\gamma$ (CAUSALIDAD MENTAL).

La imposición de determinadas constricciones de carácter semántico por parte del contexto normativo en la mente de un ser humano (la causa) hace que el proceso de selección y refuerzo de las rutas neurales y neuromusculares conducentes a la ejecución de ciertas acciones (el efecto) sea más probable que en ausencia de esas constricciones, siendo indispensable que dicho proceso finalmente se produzca.

El nuevo vínculo entre la poda sináptica y este concepto de causalidad mental consistiría en la consideración de la poda sináptica como el efecto causal en la tesis  $\gamma$  para un proceso de deliberación iterativa, lo que nos permite acotar  $\gamma$  al caso específico de otra posible modalidad de causalidad mental que nos hemos dado en llamar *poda deliberativa*:

#### Tesis $\delta'$ (PODA DELIBERATIVA).

La imposición de determinadas constricciones de carácter semántico con estructura de bucle extraño por parte del contexto normativo interno en la mente de un ser humano (la causa) hace que el proceso de poda sináptica en el que se seleccionan y refuerzan las rutas neurales conducentes a la ejecución de ciertas acciones (el efecto) sea más probable que en ausencia de esas constricciones, siendo indispensable que dicho proceso finalmente se produzca.<sup>26</sup>

Con la tesis de la poda deliberativa tratamos de integrar de un modo plausible las tres condiciones fundamentales que, desde el capítulo 4, nos hemos venido marcando para dar con una teoría libertarista y materialista del libre albedrío: la

---

<sup>26</sup> Según Alicia Juarrero (2009, p. 98), las acciones de índole intencional consisten en “el comportamiento que emana de y está determinado por constricciones simbólicas autoconscientes – por nuestro carácter excepcionalmente individualizado y simbólicamente organizado, en otras palabras”. Parece que  $\delta'$  es bastante compatible con esta idea.

causalidad mental, la correlación inteligible entre esta y un sustrato material, y, finalmente, la indeterminación ontológica que permita el control último. No obstante, creemos conveniente añadir varias observaciones referidas a estas condiciones en el ámbito de nuestra propuesta.

### 5.7. Observaciones sobre la poda deliberativa (I): causalidad mental <sup>27</sup>

En el apartado 4.1.1., estuvimos analizando cinco importantes tesis que, tal y como sostiene Carlos Moya, parecen establecer un conflicto con la causalidad mental: (a) el anomalismo en el marco de lo mental, (b) el externismo semántico e intencional, (c) el funcionamiento estrictamente sintáctico del cerebro, concebido como el órgano responsable del comportamiento intencional, (d) el cierre causal de la física, y (e) la concepción jerárquica de la realidad que parte del nivel físico como metafísicamente fundamental. Veamos qué implicaciones tiene la poda deliberativa (tesis  $\delta$ ) en relación a cada una de ellas.

La primera tesis, el anomalismo de lo mental, sostenía que no hay leyes estrictas de carácter psicofísico ni psicológico, de forma que no se puede reducir lo mental al lenguaje de la física. Según Davidson, su más relevante defensor, las relaciones de causalidad obedecen a leyes generales (concepción nomológica de la causalidad), pero las leyes que conectan causalmente lo mental con lo físico son sólo físicas. Por ello, el anomalismo no sería óbice para que las razones sean causas de las acciones y, en consecuencia, para la causalidad mental. Ahora bien, ya destacamos que este argumento

---

<sup>27</sup> Debemos aclarar que las implicaciones que aquí tratamos se refieren exclusivamente a las situaciones en que la poda deliberativa es aplicable: las deliberaciones iterativas. No entramos a valorar, por tanto, otro tipo de situaciones en las que la causalidad mental pudiera estar presente.

Davidsoniano constituye realmente una amenaza para la causalidad mental, pues conduce a una concepción epifenoménica en la que las propiedades mentales pueden servir para justificar racionalmente las acciones, pero no pueden influir causalmente sobre lo físico. Sin embargo, Moya nos brindaba una posible salida a este problema: podemos dejar a un lado la causalidad nomológica y mantener el anomalismo si planteamos una influencia causal de las razones sobre las acciones a través de una relación normativa, y no por medio de leyes científicas. Pensamos que la poda deliberativa permite adherirse a esta opción. Recordemos que, según nuestra propuesta, la posibilidad alternativa de acción que resulta vencedora en un proceso de deliberación iterativa se convierte en decisión distal en virtud de su adecuación a un estándar normativo establecido internamente: el autoconcepto o “yo” del agente. Los estados intencionales (deseos, creencias e intenciones) que interactúan para formar dicha posibilidad se convierten, así, en razones de la acción. Por lo tanto, el vínculo causal entre razones y acción obedecería a una relación normativa, lo que nos permitiría seguir apoyando la causalidad mental sin la necesidad de abandonar el anomalismo de lo mental. No obstante, seguir la solución expuesta por Moya implicaba una dificultad añadida: negar el cierre causal de la física. De esta tesis hablaremos muy pronto.

La segunda tesis, el externismo semántico e intencional, negaba que la naturaleza del significado de los signos y de los estados intencionales de un agente sea dependiente de factores únicamente internos (materiales o no). Los externistas sostenían, en oposición al internismo, que ni el significado de los signos ni los estados intencionales están constitutivamente supeditados a factores internos en exclusiva, pues también dependen constitutivamente de realidades externas al agente (de tipo físico o social), haya o no dependencia causal. Ya explicamos en el apartado 4.1.1. que, si bien el externismo parece más adecuado para dar cuenta de la dimensión semántica del contenido mental (cómo el contenido alude a la realidad en comunión con el lenguaje), las dimensiones epistemológica (cómo el conocimiento directo de sí mismo por

parte del agente le permite el control crítico de sus acciones y sus creencias, convirtiéndolo en responsable de ellas) y explicativo-causal (de qué manera puede el contenido mental conducirnos a ejecutar acciones) son más fácilmente explicables desde una perspectiva internista. El conflicto entre el externismo y la causalidad mental nació, precisamente, de la dificultad del externismo para explicar esta última dimensión: si el contenido semántico e intencional no viene dado constitutivamente por los factores internos, ¿cómo puede tener una influencia causal inteligible en el comportamiento del agente? Creemos, no obstante, que este conflicto puede suavizarse si se toma en cuenta la tesis  $\delta'$ . Veamos por qué.

Según nuestra propuesta, la distribución sináptica se encuentra en un estado de equilibrio termodinámico que es quebrantado al iniciarse la primera iteración deliberativa. La iteración, recordemos, comienza debido a la aparición de un conflicto intencional como resultado de la percepción de determinadas señales externas. A lo largo la deliberación iterativa, se produce una redistribución de botones sinápticos correlacionada con la selección y refuerzo de ciertas posibilidades alternativas en detrimento de otras. Cada posibilidad alternativa es resultado de la interacción entre estados intencionales diversos. Finalizada la deliberación, se obtiene la distribución sináptica  $d_i$ , distinta a la inicial ( $d_0$ ). En ese momento, los estados intencionales cuya interacción producía la decisión (o sea, la  $P$  ganadora) se han convertido en razones de la acción. Según esta explicación, el contenido intencional sufre un cambio (se pasa de un conflicto a una decisión), y a este se llega, en parte, por la modificación de uno de los factores internos que lo determinan (en concreto, la distribución sináptica), pero dicha modificación (consistente en la ruptura de un equilibrio termodinámico y el tránsito hasta un nuevo equilibrio) no habría sido posible sin la participación de factores externos (en este caso, las señales cuya percepción desencadena el conflicto intencional). La poda deliberativa constituye, por tanto, una explicación de corte externista, y, sin embargo, no impide que el contenido intencional influya causalmente en el comportamiento:

llegado el momento oportuno, la red neural correspondiente a la decisión distal, equivalente a la *P* que ha resultado vencedora por su adecuación a constricciones semánticas internas (el autoconcepto), se comunica con estructuras cerebrales responsables de las funciones motoras, que desencadenan la acción.

La tercera tesis consistía en el funcionamiento estrictamente sintáctico del cerebro, concebido este como el órgano responsable del comportamiento intencional. Según esta tesis, “los procesos y estructuras cerebrales sólo pueden ser sensibles a propiedades físicas o, a lo sumo, sintácticas de signos y eventos, no a sus propiedades semánticas, como el contenido o el significado” (Moya 2011, p. 190). Así, el contenido semántico no podría influir causalmente en el comportamiento. En otras palabras, no sería posible la causalidad intencional. Al igual que un ordenador no responde al contenido semántico, el cerebro, que también es un sistema físico, no estaría habilitado para responder al significado. Consideramos que esta postura es altamente problemática. Creemos que la equiparación entre un ordenador y un cerebro humano no es adecuada. En efecto, ambos constituyen sistemas físicos, pero las diferencias son notables. En primer lugar, la de complejidad del cerebro es infinitamente mayor y responde a una estructura jerárquica de varios niveles: átomos, moléculas, orgánulos celulares, células, subdivisiones del córtex y, finalmente, el órgano en su conjunto. En segundo lugar, el cerebro humano es capaz de responder a las emociones, algo inimaginable para un ordenador. No sólo eso, sino que las emociones parecen jugar un papel crucial en el funcionamiento de nuestro “ordenador biológico”, tal y como propone la hipótesis de los marcadores somáticos de Damasio (que expusimos en el apartado 4.1.3.). Pero la tercera diferencia entre un ordenador y un cerebro humano es, sin duda, la más reseñable. Se trata de la neuroplasticidad. Mientras que un ordenador posee un hardware inmodificable, la estructura física del cerebro, es decir, su estructura neural, puede cambiar por motivos diversos. Uno de ellos consiste en la adaptación del individuo a su entorno. Cuando expusimos la poda semántica

seguimos precisamente esta línea, pues planteamos la influencia del ambiente normativo del individuo, semánticamente estructurado, sobre la poda sináptica cerebral. De manera análoga, al proponer la tesis  $\delta'$  concebimos una influencia normativa y semántica similar, pero de origen contextual interno, sobre la poda sináptica. Aceptar esta propuesta supondría negar con rotundidad la tesis del funcionamiento estrictamente sintáctico del cerebro.

La cuarta tesis, el cierre causal de la física, defendía que “todo cambio físico tiene una explicación física completa” (Moya 2006a, p. 209). Admitir que la mente ejerce una influencia causal sobre la conducta gracias a sus propiedades específicamente mentales constituye una negación a esta tesis en tanto en cuanto supone afirmar que no existe una explicación física completa para eventos que son de carácter físico. No obstante, esta es la línea argumental a la que conduce irremediamente la poda deliberativa. Según  $\delta'$ , tiene lugar una influencia causal, por medio de constricciones internas, de un estándar semántico (el autoconcepto o “yo” del agente) sobre el proceso de poda sináptica. En consecuencia, si admitimos que la poda sináptica es un proceso físico, la poda deliberativa postula que, al menos en el ámbito de una deliberación deliberativa, la explicación de dicho proceso no puede efectuarse íntegramente en términos físicos, sino que requiere acudir a realidades mentales (las constricciones semánticas) para ser completada.

La tesis del cierre causal de la física va indisolublemente unida a la quinta de las tesis que mencionábamos al principio de la sección. Se trata de la concepción jerárquica de la realidad que parte del nivel físico como metafísicamente fundamental: si pensamos en el universo como una realidad jerárquicamente estructurada y si atendemos a un criterio de orden temporal y de causalidad, los niveles físicos constituirían las capas metafísicamente fundamentales del universo, de manera que la física sería capaz de explicar en último término la eficacia causal de los niveles estudiados por el resto de ciencias, siendo imposible lo contrario. Esta tesis implica un sentido ascendente de la causalidad entre los niveles jerárquicos, en oposición a la

causalidad mental, que procede de forma descendente. La poda deliberativa implica negar que el nivel físico sea metafísicamente fundamental, al menos en el sentido en que acabamos de plantear este postulado.  $\delta'$  asume un reduccionismo ontológico débil, conforme al cual no hace falta que surjan nuevos componentes de carácter ontológico conforme ascendemos por la jerarquía de la realidad (véase apartado 4.2.2.). Es decir, da por sentado que todas las piezas básicas sobre las que se construye la realidad son de naturaleza física. Ahora bien, este planteamiento no empuja a defender un reduccionismo fuerte, para el que sólo son ontológicamente reales los elementos que pertenecen al nivel inferior, siendo el resto meros agregados perecederos de estos. Esta clase de reduccionismo es la que se esconde tras la tesis que estamos tratando. Ahora bien, según nuestra propuesta, las constricciones semánticas internas (de naturaleza mental) son capaces de inducir cambios en la estructura neural de un individuo por medio de la poda sináptica. Si se diera por buena esta proposición, se estaría aceptando una influencia causal de un determinado nivel jerárquico, el mental, sobre otro de orden inferior, el fisiológico. Desde luego, resulta problemático concebir que un elemento jerárquico capaz de provocar alteraciones en otros niveles no goce de un estatus ontológico propio. Es por ello que la poda deliberativa resulta incompatible con el reduccionismo ontológico fuerte y, en consecuencia, con la tesis según la cual el nivel físico es el metafísicamente fundamental en la jerarquía de la realidad.

Acabamos de indicar las implicaciones que la tesis  $\delta'$  tiene para las cinco tesis que mencionábamos al principio, pero no querríamos finalizar la sección sin comentar también las consecuencias de nuestra propuesta en relación a dos importantes asuntos también relacionados con la causalidad mental, y analizados en la sección 4.1. Se trata del vínculo entre razones y acción y del papel de los estados fenomenológicos en la causalidad mental.

En relación al primer asunto, ya vimos que existen teorías tanto causales como no causales de la acción intencional. Las teorías no causales

defendían que las razones pueden justificar las acciones pero no causarlas. En cuanto a las teorías causales, defienden que las razones pueden constituir causas de las acciones. Entre este segundo grupo de teorías destacaba la desarrollada por Davidson. Sin embargo, esta era vulnerable a dos amenazas: el epifenomenismo y las cadenas causales desviadas. Recordemos que, según proponía Moya, este problema podría subsanarse si no se conciben la justificación y la causalidad como independientes entre sí, sino que se fusionan de tal forma que la justificación de una acción intencional es al mismo tiempo su causa:

El núcleo esencial de nuestra propuesta es el siguiente: si una razón no causa una acción, la acción en cuestión no está justificada por dicha razón. Así, la única condición que se requiere para que una acción sea intencional es que dicha acción esté justificada, bajo la descripción adecuada, por las razones del agente. El requisito de causalidad ya está incluido en la condición de justificación (Moya 2006a, p. 204).

Pensamos que la poda deliberativa es compatible con este planteamiento. Recordemos cómo se resuelve el conflicto intencional en nuestra propuesta. Durante las sucesivas iteraciones deliberativas, las distintas posibilidades alternativas ven modificadas sus probabilidades de selección. En el momento en el que concluye la iteración  $i$ , una única  $P$  ha resultado vencedora del conflicto, por lo que su probabilidad de selección es igual a 1. De esta manera, se convierte en la decisión distal y, cuando se presente el momento apropiado, dará como resultado la acción vinculada a ella.<sup>28</sup> Los deseos, creencias e intenciones cuya interacción ha conformado la decisión se convierten, así, en las razones de la acción. Pues bien, es sólo en ese preciso instante cuando las razones pueden, *en principio*, justificar la acción y, al mismo tiempo, causarla. Dado que, justo con anterioridad, la  $P$  a la que dan lugar no tiene una

---

<sup>28</sup> Estamos pensando en una situación ideal. La decisión podría verse incumplida porque intervengan factores ajenos por completo al control del agente. Por ejemplo, este podría morir inesperadamente antes de que se dé la oportunidad de ejecutar la acción vinculada.

probabilidad absoluta de ser la finalmente seleccionada, ninguna  $P$  se encuentra aún en disposición de explicar ni causar la acción que tendrá lugar más adelante. Cuando una de ellas alcance una probabilidad total, podrá, *en principio*, justificar y a la vez causar la acción. Se consigue así cumplir con un requisito demandado por Morden para que nuestra conducta pueda ser considerada como libre: que la cadena causal atraviese al agente en lugar de rodearlo (véase sección 5.5.). Hemos enfatizado “en principio” porque la acción sólo será intencional si, como bien recalca Moya, es ejecutada conforme a la descripción apropiada, de manera que evitemos la amenaza de las cadenas causales desviadas. Nuestra tesis  $\delta'$  da por sentado este requisito, por lo que no permite esquivar la amenaza, aunque, como ya hemos dicho, sí es compatible con la fusión de justificación y causalidad propuesta por Moya.

En relación al papel de los estados fenomenológicos en la causalidad mental, vimos que la convergencia entre estos y los estados intencionales puede dar lugar a estados mixtos, es decir, a emociones y sentimientos. En ellos hay una manifestación fenomenológica que está dirigida a un contenido intencional. También pudimos ver que la hipótesis de los marcadores somáticos marcaba una vía muy interesante para profundizar en el mecanismo de aparición de dichos estados. El papel de los marcadores somáticos en la poda deliberativa resulta clave, ya que actúan como constricciones que, al inicio de cada iteración deliberativa, permiten seleccionar rápidamente ciertas posibilidades alternativas en detrimento de otras, haciendo así de la deliberación un mecanismo más simple y eficaz. En nuestra propuesta, por tanto, la causalidad mental está presente no por medio de estados intencionales en exclusiva, sino a través de un intrincado mecanismo de interacción entre estos y ciertos estados fenomenológicos.

## 5.8. Observaciones sobre la poda deliberativa (II): correlación inteligible entre la causalidad mental y un sustrato material

En la sección 4.2. apuntábamos que el funcionamiento de los sistemas complejos les permite albergar procesos de causalidad descendente en virtud de un vínculo de emergencia contextual entre las propiedades de orden superior y las propiedades de niveles inferiores. También sosteníamos que, en ocasiones, dicha causalidad se hace efectiva en virtud de constricciones normativas que, mediante su conexión con los niveles mental y biológico, ejercen una causalidad estructurante dretskeana sobre componentes físicos. La relación entre mente y materia así concebida nos sirvió para plantear la tesis  $\gamma$  (causalidad mental), sobre la cual se sustenta  $\delta'$ . Ahora nos gustaría añadir algunas observaciones sobre el papel de la causalidad en la poda deliberativa.

La tesis  $\delta'$ , tal y como sucedía con sus predecesoras  $\gamma$  y  $\delta$ , puede ser vista como una versión de la propuesta de Dretske para la naturalización del contenido mental e intencional: determinadas conexiones neurales y neuromusculares son *reclutadas*, mediante poda sináptica, para ejercer funciones concretas debido a aquello que *indican* sobre los signos presentes en el contexto normativo. En este sentido, podríamos considerar que la causalidad descendente actuante en la poda deliberativa actúa a modo de causalidad estructurante dretskeana. Ahora bien, según la propuesta de Dretske, recordemos, se produce una conexión entre los estados internos y el entorno por medio de un vínculo causal de tipo nomológico establecido por aprendizaje (véase apartado 4.2.1.). De aquí surgen dos discrepancias reseñables en comparación con nuestra propuesta. En primer lugar, Dretske piensa en un contexto normativo de índole ambiental, basado en leyes naturales. En la poda deliberativa, en cambio, el contexto normativo es el autoconcepto del agente, de tipo interno. En segundo lugar, nuestra propuesta concibe un vínculo entre causa y efecto que, a diferencia del dretskeano, no es

de carácter nomológico. Desarrollaremos a continuación esta segunda discrepancia.

Según vimos en el capítulo 4, Moya proponía ver la causalidad como una relación en la cual un evento A (la causa) hace que la aparición de otro evento B (el efecto) sea más probable que cuando A no tiene lugar, bajo el imperativo de que B, de hecho, suceda. Dicha noción, que hemos empleado como tesis preliminar  $\alpha$  (sobre cuyos cimientos hemos desarrollado  $\beta$ ,  $\gamma$ ,  $\delta$  y  $\delta'$ ), podía encajar, como opinaba el propio Moya, “con distintas visiones de la causalidad, deterministas o probabilistas, nomológicas, contrafácticas, y otras” (Moya 2011, p. 187). Sin embargo, la tesis  $\delta'$  no se basa en una noción nomológica (humeana) de la causalidad. Conforme a esta, el vínculo entre causa y efecto se da por necesidad y en virtud de leyes naturales. Dicho de otro modo: las leyes naturales *obligan* a que el evento que constituye la causa desemboque en el evento que constituye el efecto. Pero este no es el caso de la poda deliberativa. En nuestra propuesta, la causa (la imposición de determinadas constricciones semánticas internas) no siempre desemboca en el efecto (la poda sináptica en la que se seleccionan y refuerzan las rutas neurales que conducen a ejecutar ciertas acciones). Recordemos que en los interciclos deliberativos (intervalos temporales en los que las constricciones del “yo” dejan de intervenir provisionalmente) hay una evolución aleatoria de la distribución sináptica, lo que provoca que las probabilidades de cada posibilidad alternativa  $P$  puedan cambiar o que surjan nuevas  $P$ . De esta manera, cabe la opción de que una misma imposición inicial de constricciones semánticas por parte del autoconcepto del agente (una misma causa, según el concepto de esta propuesto en  $\delta'$ ) conduzcan a procesos de poda sináptica en los que se seleccionen y refuercen rutas neurales conducentes a la ejecución de acciones diversas (distintos efectos, según el concepto de estos propuesto en  $\delta$ ). Además, la poda sináptica no responde a una normatividad basada en leyes naturales, sino a una de carácter semántico: el “yo” o concepto que de sí mismo guarda el individuo. Por tanto, en la poda deliberativa hay un vínculo

causal que, si bien es de tipo normativo, no es de tipo nomológico, pues la causa: 1) no obliga a que se dé el efecto, y 2) no conduce a este en virtud de leyes naturales. La causalidad operante en nuestra propuesta no se ajusta a la noción humeana de la causalidad, y sí encaja, en cambio, con la siguiente descripción:

Algunos filósofos [...] <sup>29</sup> han sostenido que Hume está exagerando un tanto la fuerza de la conexión que debe mantenerse entre la causa y el efecto. Parece desde luego posible que *A* haya provocado *B* aun dada la suposición de que no era *imposible* que *A* ocurriera sin estar seguida por *B*. Un comentario provocador de un hablante, por ejemplo, podría ocasionar que el oyente respondiera airadamente, aun siendo posible que el oyente hubiera reprimido la reacción emocional. Sin embargo, Hume tiene verdaderamente razón en que una conexión causal implica algún tipo de conexión no accidental entre la causa y el efecto. La existencia de la causa generalmente incrementa la probabilidad de la existencia del [efecto]. Además, en casos normales, la causa satisface una condición necesaria para el efecto, en el sentido de que el efecto no podría haber ocurrido *en esas mismas circunstancias* [la cursiva es nuestra] en ausencia de la causa. Un mundo completamente desprovisto de todas esas conexiones no accidentales sería un mundo sin causalidad (Koons y Pickavance 2015, p. 42).

Con la poda deliberativa proponemos que, en relación a un vínculo entre la imposición de determinadas constricciones normativas internas (causa) y la poda sináptica en virtud de la cual se seleccionan y refuerzan ciertas rutas neurales (efecto), si no se impusieran esas constricciones, entonces no podría suceder esa poda *en las mismas circunstancias*, esto es, cuando el resto de factores relevantes para el vínculo causal participaran de idéntica manera. Por tanto, la imposición de las constricciones internas constituye una condición necesaria para que se dé la poda. Ahora bien, las constricciones no son suficientes para desencadenarla. Obviamente, resulta imprescindible el concurso de componentes físicos, químicos y biológicos: átomos, biomoléculas, neuronas, cerebro, etc. También deben participar factores fisiológicos: los impulsos sinápticos y los procesos de percepción. Finalmente, se requiere la intervención

---

<sup>29</sup> Los autores citan aquí a Elizabeth Anscombe (1975).

de estados fenomenológicos, que actúan por medio de los marcadores somáticos. En definitiva, la poda deliberativa implica la *cooperación causal* entre causas necesarias cuyos sentidos de influencia causal (ascendente, intranivel y descendente) se complementan entre sí (cf. Ellis 2009, pp. 76-9; cf. Bishop y Atmanspacher 2011, p. 109). Dicha cooperación es compatible, desde luego, con la fórmula propuesta por Morden (1990, p. 72; véase sección 5.5.): “Soy (suficientemente) libre cuando soy una de las causas de mi propia conducta”.

Hasta el momento, nuestro análisis ha tenido que ver con una perspectiva *espacial* de la causalidad. Ahora bien, ¿cuál es la relación *temporal* entre causa y efecto en la tesis  $\delta$ ? Según Bishop y Atmanspacher (2011, p. 109), “la causalidad en el sentido de causalidad eficiente sólo es aplicable a las relaciones horizontales (i.e., diacrónicas) entre estados del sistema en el mismo nivel descriptivo. Para las relaciones verticales (i.e., sincrónicas) entre niveles de descripción no tiene sentido hablar de una secuencia causal temporal”. Si hacemos caso a esta opinión, una defensa de la causalidad descendente nos obligaría a argumentar a favor de la simultaneidad entre causa y efecto, Pero, ¿cómo podemos saber si existe una relación causal entre un evento A y un evento B si ambos suceden a la vez? ¿No serán, quizá, dos manifestaciones idénticas, aunque descritas a distinto nivel, de un mismo efecto causal provocado por un tercer evento anteriormente acaecido? Y, suponiendo que realmente exista una relación causal entre A y B, ¿cómo distinguir si A causa B o si es a la inversa? Como vemos, y tal y como sostiene Kim (1999, pp. 28-9), la inteligibilidad de la causalidad descendente resulta difícil si la concebimos como sincrónica. Aunque la sincronía causa-efecto tiene sus defensores (e.g.: Huemer y Kovitz 2003; Bishop 2008, pp. 242-4), la relación temporal que entre ambos eventos se establece en la poda deliberativa es de una naturaleza distinta. En contra de la opinión de Bishop y Atmanspacher citada más arriba, estamos pensando en un vínculo que, pese a su verticalidad, se presenta secuencialmente. En otras palabras, se trataría de una relación diacrónica, en la

que la causa antecede al efecto. Pero, además, sería una relación continua, cuya comprensión, en palabras de Koons y Pickavance (2015, pp. 200-201),

radica en la noción de un *proceso* extenso temporalmente. Suponga, por ejemplo, que el agente *A* ejerce la capacidad de mover *P*. ¿Qué ocurre en el momento *t* de la acción? El agente *A* ejerce su capacidad en *t*, y en ese preciso momento [...] *P* inicia un proceso de movimiento. Sin embargo, el proceso se extiende necesariamente hacia el futuro, al menos hasta que sea interrumpido por alguna acción más. Las etapas anteriores de este proceso transmiten la influencia causal de *A* sobre *P* hacia sus etapas posteriores. Esta transmisión es continua y densa en vez de discreta, ya que entre cualquier etapa posterior del proceso y cualquier etapa anterior existe un número infinito de etapas instantáneas, cada una causalmente previa a la una y causalmente posterior a la otra.

Según los autores,

la causalidad continua implica un *continuum* [cursiva añadida] indiviso de eventos. En el supuesto básico, dos eventos están conectados por causalidad continua cuando ambos son partes de un único proceso. Un proceso real es un todo extenso temporalmente que es más fundamental metafísicamente que cualquiera de sus inextensas e instantáneas partes. Las partes posteriores del proceso dependen de las partes anteriores porque ambas son partes del mismo proceso, no porque exista alguna conexión discreta o cadena de conexiones discretas entre las dos (*ibid.*, p. 201).

Si aceptamos la tesis  $\delta'$ , la imposición de ciertas constricciones internas de carácter semántico (la causa) provoca, durante una iteración deliberativa, que tenga lugar un proceso de poda sináptica en determinadas rutas neurales (el efecto). En armonía con la descripción anterior, podemos decir que la poda sináptica comienza cuando aparecen las constricciones. En ese momento empiezan a retraerse ramales axónicos y a eliminarse espinas dendríticas en las rutas neurales descartadas semánticamente, y también empiezan a reforzarse, mediante la creación de nuevas sinapsis, los vínculos interneuronales pertenecientes a las rutas seleccionadas. Ahora bien, mientras las constricciones sigan presentes, estos fenómenos biológicos comunicarán la relación causal a los fenómenos siguientes (*i.e.*: ulteriores retracciones, eliminaciones y refuerzos), y el proceso finalizará cuando llegue un interciclo o cuando se haya

resuelto el conflicto intencional. En cada iteración, la relación entre causa y efecto es de carácter *continuo*, porque no es posible asignar una extensión temporal que no resulte arbitraria a las etapas intermedias: en todo momento hay vínculos que se debilitan a la par que otros se refuerzan. Lo único que podríamos hacer es describir la distribución sináptica correspondiente a un instante infinitesimal cualquiera. Todas estas etapas (inicial, intermedias y final) constituyen, pues, partes de un único proceso biológico: la poda sináptica. La relación causal es, además, de carácter secuencial o *diacrónico*. En la explicación de Koons y Pickavance, el agente  $A$  hace uso de su facultad para mover  $P$  en cierto momento  $t$ , y es entonces cuando  $P$  se mueve en un proceso extendido hacia el futuro. La causa, aquí, antecede al efecto. Del mismo modo, podemos decir que, en la poda deliberativa, la imposición de constricciones en una iteración deliberativa desencadena en cierto momento la poda sináptica, y es entonces cuando esta se desarrolla en un proceso extendido hacia el futuro.

Finalizamos ahora la presente sección. He aquí la principal conclusión que podemos extraer: la poda semántica implica una causalidad que no es sólo *descendente, mental* y, aunque con matices, *estructurante* al modo dretskeano, sino que también es *diacrónica* (secuencial) y *continua*. Todo ello, recordemos, en el seno de un sistema no integrable (el cerebro) que está englobado en un sistema complejo (el ser humano), y en virtud de una relación de emergencia contextual entre las propiedades mentales y aquellas propias de niveles jerárquicos inferiores.

### **5.9. Observaciones sobre la poda deliberativa (III): indeterminación ontológica que permita el control último**

La causalidad mental y su adecuada correlación con componentes y procesos físicos se antojan imprescindibles para cualquier postura respecto al libre

albedrío desde que aspire a estar en armonía con la ciencia actual. Sin embargo, el componente que diferencia al libertarismo es la apuesta por el indeterminismo y, en concreto, por aquel que sea de carácter ontológico y que permita el control último del agente respecto a sus decisiones y/o acciones.

Al principio de este trabajo definíamos el determinismo de tres maneras distintas. Según Fischer *et al.* (2007, p. 2), “podemos tratar el determinismo como la tesis según la cual en cualquier instante [...] el universo posee exactamente un solo futuro posible físicamente”. Por su parte, van Inwagen (1983, p. 65), sostiene que el determinismo es la conjunción de dos tesis: (1) “Para cada instante de tiempo, existe una proposición que expresa el estado del mundo en ese instante”, y (2) “[s]i  $p$  y  $q$  son proposiciones cualesquiera que expresan el estado del mundo en algunos instantes, entonces la conjunción de  $p$  junto con las leyes de la naturaleza implica  $q$ ”. O, dicho de otro modo, “[e]l determinismo es, intuitivamente, la tesis según la cual, dados el pasado y las leyes de la naturaleza, existe sólo un futuro posible” (*id.*). En los tres casos, la idea esencial es la misma: teniendo en cuenta las condiciones impuestas por la naturaleza en cierto momento, los acontecimientos futuros no pueden sino desarrollarse de un único modo. El indeterminismo niega la validez universal de esta idea, es decir, se opone a que el determinismo sea cierto en todo momento y lugar. Ahora bien, como ya expusimos en el apartado 2.1.7., el indeterminismo puede ser de índole epistemológica u ontológica. La primera variedad consiste en una mera impredecibilidad debida a nuestra incapacidad para conocer cómo se comportará un sistema. La variedad ontológica, por su parte, constituye un aspecto esencial del sistema estudiado, cuya impredecibilidad, por tanto, se presenta *per se* y no se debe a nuestras limitaciones observacionales. Es esta segunda clase de indeterminismo la que resulta imprescindible para que el agente pueda gozar de control último sobre sus decisiones y/o acciones.

La poda deliberativa defiende una posición que puede ser calificada como ontológicamente indeterminista en tres aspectos distintos pero complementarios entre sí:

(1) *Dimensión subjetiva.* Mientras el conflicto intencional permanece abierto, el agente desconoce, desde su perspectiva interna, cómo se resolverá este. Percibe, así, que goza de posibilidades alternativas. Esta incertidumbre ante su propio proceder futuro motiva al agente para continuar deliberando, lo que constituye una influencia sobre la evolución futura del cerebro, dado que la poda deliberativa reconfigura la estructura neural. Dicha influencia sobre componentes físicos del sistema estudiado permite sostener que la impredecibilidad se nos presenta como un componente esencial e intrínseco de este.

(2) *Dimensión física.* La evolución de la distribución sináptica entre  $d_0$  y  $d_t$ , (correspondientes a distintas fases termodinámicas) posee dos importantes características: sólo puede ser estudiada estadísticamente y, además, es irreducible al comportamiento individual de las neuronas. Si consideramos el cerebro como un sistema no integrable, entonces, en línea con los planteamientos del Grupo Bruselas-Austin (véase sección 5.4.), cabe decir que las distribuciones de probabilidad que surgen constituyen una propiedad esencial e intrínseca de dicho sistema. Y si consideramos válida la distribución de botones sinápticos como modelo descriptivo del sistema, entonces, en línea con la sugerencia de Bishop (véase nuevamente sección 5.4.), cabe decir que entre dichas distribuciones y las neuronas individuales se establece una relación de emergencia contextual: la descripción del comportamiento de las neuronas individuales es necesaria pero insuficiente para explicar la evolución de la distribución. Ambas características nos indican que la conjunción de la disposición espacial de los botones sinápticos en un momento dado y las leyes de la física subyacentes a los mecanismos celulares actuantes en cada neurona no determina de qué manera se redistribuirán dichos botones en el futuro (y tampoco si lo harán). En cambio:

Las leyes fundamentales de la física definen un espacio de posibilidades físicas, en otras palabras, definen los comportamientos físicos posibles de acuerdo con las leyes de la física. Esto contrasta con aquellas posibilidades que son posibles lógicamente pero violarían las leyes fundamentales (e.g., elevarme levitando contra la fuerza de la gravedad mediante la pura intención), lo cual constituye un espacio mucho mayor. Ejemplificaciones concretas de las condiciones iniciales y de contorno, junto con los contextos en los que las leyes y condiciones vienen a expresarse, constriñen luego más el espacio de posibilidades físicas. En ausencia, entonces, de cualesquiera otras constricciones contextuales y propiedades, la física fundamental proporciona las condiciones que definen el espacio de posibilidades para el comportamiento y las interacciones de la materia (Bishop 2012, p. 68).

Nuestra idea es que, a la vez que vetan las opciones incompatibles con ellas, las leyes que gobiernan el dominio físico permiten un abanico de alternativas para la evolución futura de la distribución sináptica, cuyo transcurso se verá decisivamente influido por el contexto interno. No obstante, esta influencia puede manifestarse de dos modos distintos, que establecen dos variedades de la dimensión física del indeterminismo en nuestra propuesta. Hablamos de una dimensión normativa y de otra aleatoria.

(2a) *Dimensión normativa.* A lo largo de las sucesivas iteraciones deliberativas, el componente normativo, no sujeto a leyes naturales, resulta decisivo para conducir los acontecimientos en una u otra dirección entre las diversas posibles. También es importante el componente fenomenológico presente en los marcadores somáticos.

(2b) *Dimensión aleatoria.* En los interciclos cambia el contexto interno, ya que las constricciones normativas y fenomenológicas relativas al conflicto intencional cesan en su influencia. En estos intervalos temporales, la deriva sináptica provoca que se formen correlaciones neuronales impredecibles y que la distribución sináptica muestre efectos colectivos de carácter aleatorio.

(3) *Dimensión cíclica.* La influencia de factores aleatorios durante un interciclo hace que, dada una cierta distribución sináptica al final de la iteración previa, sean posibles diversas distribuciones en el momento de iniciarse la

iteración siguiente. De este modo, el número de iteraciones (ciclos) necesarias para zanjarse el conflicto intencional se torna intrínsecamente impredecible.

Las dimensiones que acabamos de describir interactúan para dar lugar al indeterminismo de nuestra propuesta. Por un lado, la dimensión subjetiva actúa como factor motivacional que mantiene viva la deliberación y, con ella, el proceso físico subyacente. La dimensión física, por su parte, hace que la evolución de la distribución sináptica sea posible de varias maneras distintas. Finalmente, la combinación de los factores normativo y aleatorio de la dimensión física provoca la aparición de una dimensión cíclica, según la cual el número de iteraciones que se requieren para solucionar el conflicto intencional es del todo incierto. Todas estas dimensiones conducen a una conclusión conjunta: teniendo en cuenta las condiciones impuestas por la naturaleza en cierto momento, los acontecimientos futuros pueden desarrollarse de varios modos (indeterminismo).

Los libertaristas, tal y como ya vimos en la sección 1.3., sostienen que las posibilidades alternativas y el control último son necesarios para la libertad y que el determinismo excluye ambos. Por otro lado, y conforme a lo que vimos en la sección 1.1., diremos que:

(1) un agente dispone de *posibilidades alternativas* cuando goza de un número, mayor que uno, de posibilidades entre las cuales escoger en sus decisiones y/o acciones.

(2) ese mismo agente tiene *control último* si es la fuente, el autor, el origen último de sus decisiones y/o acciones.

Si aceptamos la validez del indeterminismo que hemos propuesto para la poda deliberativa, resulta sencillo aceptar la primera condición. Al comenzar la deliberación iterativa, el individuo dispone de diferentes posibilidades de elección, a las que bautizamos con la letra *P*. Dado que, al comenzar la deliberación, los acontecimientos futuros pueden darse de diferentes maneras (indeterminismo), resulta inteligible que cualquiera de dichas posibilidades tenga opciones reales de resultar vencedora en el conflicto intencional.

En cuanto a la segunda condición, el control último, podríamos aducir dos motivos para defender que el agente controla la decisión distal adoptada al final de la deliberación iterativa. En primer lugar, a dicha decisión se llega tras un proceso en el que las posibilidades alternativas son confrontadas al “yo” o concepto que de sí mismo tiene el individuo, por lo que este puede ser considerado su autor. En segundo lugar, la acción que es fruto de la decisión distal está causada y justificada por razones. Parece ser, por tanto, que la búsqueda del control de la decisión distal nos conduce al agente. Además, si aceptamos el indeterminismo propuesto, cabe sostener que los acontecimientos anteriores al comienzo de la deliberación no determinan la decisión, por lo que el origen último de esta se halla en el individuo deliberante. Daría la sensación, por tanto, de que se satisface la condición de control último. Ahora bien, muchos filósofos no encontrarían esta explicación suficiente, exigiendo que el agente, además de ser responsable de sus acciones, también lo fuera de sí mismo y de los motivos que le llevan a ejecutarla. ¿Es posible satisfacer este requisito en nuestra propuesta?

A nuestro modo de ver, el rol normativo desempeñado por el autoconcepto o “yo” del agente en la poda deliberativa es similar al que ejercen las creencias evaluativas en la propuesta libertarista de Moya. Según este,

los juicios prácticos sobre qué acción es mejor, o mejor que las alternativas, deberían ser entendidos como la aplicación de las creencias evaluativas de un agente, como estándares normativos, a una situación particular a la que se enfrenta. Un juicio práctico es un fenómeno cognitivo más que volitivo. No es propiamente una elección, sino, por así decirlo, la expresión de una creencia sobre el valor de modos de actuación particulares, formados a la luz de visiones evaluativas más generales. Por regla general, las elecciones racionales de un agente son fijadas por sus juicios prácticos. Las elecciones no pueden ir desprovistas de juicios prácticos sin caer en la arbitrariedad. En nuestra aproximación cognitiva, entonces, los juicios prácticos sustituyen a las elecciones como pasos centrales en la deliberación práctica. Las elecciones, a su vez, dependen de los juicios prácticos y permanecen vinculadas a ellos. A un nivel más profundo, que supone la base del control último del agente sobre sus elecciones y acciones, tenemos de nuevo, por supuesto, creencias – de tipo evaluativo – más que elecciones radicales (e.g., las voluntades autoformadoras de Kane). Estas creencias se forman, a su vez, a la luz de otro

conjunto de estándares normativos, como la coherencia y la sensibilidad hacia los hechos (Moya 2006b, pp. 203-4).

En analogía con esta explicación, diremos que las decisiones tomadas por un agente tras una deliberación iterativa dependen de un juicio práctico que, en última instancia, está basado en su “yo” o autoconcepto, el cual alberga las creencias evaluativas del agente sobre sí mismo. Estas y el autoconcepto se forman gracias a la sensibilidad y la coherencia del agente hacia las consecuencias de sus acciones. En este sentido, el agente gozaría de responsabilidad y control de tipo cognitivo (no volitivo) sobre las mencionadas creencias, que intervienen como motivos para sus decisiones y acciones, y sobre sí misma (*i.e.*, sobre su “yo”). En virtud de la analogía expuesta, pensamos que es extensible a nuestra tesis  $\delta'$  la desvinculación que Moya efectúa entre el control último y elecciones últimas (o “radicales”), como las voluntades autoformadoras, o SFWs (“self-forming willings”), propuestas por Kane. También cabe aplicar la desvinculación con respecto a las SFAs (“self-forming actions”), que siguen a las SFWs y que, en el apartado 4.3.1., definíamos como acciones o elecciones pasadas que conforman el carácter que el agente tiene en la actualidad. A ambas, SFWs y SFAs, se les han atribuido problemas de regresión infinita (*e.g.*, Moya 2006b, pp. 154-5).

Para continuar con la presente sección, no gustaría recordar que existe un argumento según el cual el indeterminismo excluye cualquier tipo de control del agente sobre sus decisiones y/o acciones porque el indeterminismo conlleva el azar y no existe control sobre algo que es azaroso. Este argumento, que ya avanzamos en la sección 1.3., es denominado “argumento de *Mind*” u “objeción de la suerte”, y constituye una gran amenaza para la inteligibilidad del libertarismo. Existen distintas versiones del argumento. Una de las más citadas en los últimos años, conocida como “Rolling-Back” (Moya 2006b, p. 136), ha sido propuesta por van Inwagen (2000, pp. 13-16), quien nos invita a imaginar una marcha atrás en el tiempo:

Permitámonos suponer que los actos libres indeterminados ocurren. Supongamos, por ejemplo, que en cierta situación difícil Alice se enfrentara a una elección entre mentir y decir la verdad y que libremente escogiera decir la verdad [...]. Y permitámonos asumir que el libre albedrío sea incompatible con el determinismo, y que el hecho de que Alice diga la verdad, al ser un acto libre, esté por tanto indeterminado. Ahora supongamos que, inmediatamente después de que Alice dijera la verdad, Dios provocase que el universo volviera exactamente a su estado un minuto antes [...]. ¿Habría mentido o habría dicho la verdad? Dado que la decisión “original” de Alice, su decisión de decir la verdad, estaba indeterminada—ya que estaba indeterminado si mentiría o diría la verdad—, su “segunda” decisión estaría también indeterminada (*ibid.*, p. 14).

Seguidamente, van Inwagen propone que imaginemos que la operación de “rebobinado” hasta el estado del universo existente un minuto antes de la primera decisión de Alice tuviera lugar en mil ocasiones, y sostiene:

[D]e nuevo, no podemos decir qué habría sucedido, pero podemos decir lo que *probablemente* habría sucedido: a veces Alice habría mentido y a veces habría dicho la verdad. A medida que aumente el número de “reproducciones”, los observadores advertiremos—casi con certeza—que el ratio entre el resultado “verdad” y el resultado “mentira” se estabiliza sobre, converge en, cierto valor. [...] [P]ermitámonos imaginar el caso más simple: observamos que Alice dice la verdad en aproximadamente la mitad de las reproducciones y miente en aproximadamente la mitad de las reproducciones. [...] ¿No es cierto que, a medida que observemos que el número de reproducciones aumenta, nos convenceremos de que lo que ocurrirá en la *siguiente* reproducción es una cuestión de azar? (*ibid.*, pp. 14-15).

Según el autor, este razonamiento es igualmente válido para el primero de todos los procesos que conducen a una acción de Alice (aunque este no consiste en una reproducción propiamente dicha), y lo es también en el caso de que Dios no revierta el tiempo, que es lo que sucede en realidad. La conclusión es que

una acción indeterminada es simplemente una cuestión de azar: si estaba indeterminado, en el caso único y real, que Alice mintiera o dijera la verdad, era una mera cuestión de azar que mintiera o dijera la verdad. Si supiéramos de antemano que las probabilidades objetivas [...] de que Alice diga la verdad y de que Alice mienta eran ambas 0.5, entonces (suponiendo que nuestro bienestar dependiese de que dijera la verdad) sólo podríamos considerarnos *afortunados* cuando, eventualmente, dijera la verdad. ¿Pero entonces cómo podemos decir que el hecho de que Alice diga la verdad era un acto libre? Si

se enfrentase a decir la verdad o mentir, y fuera una mera cuestión de azar cuál de estas cosas haría, ¿cómo podemos decir que—y esto es indispensable para que el acto sea libre—era *capaz* de decir la verdad y *capaz* de mentir? (*ibid.*, pp. 15-16).

Una vez expuesto el ejemplo de van Inwagen, nos disponemos a relacionarlo con el indeterminismo de la poda deliberativa. Con este objetivo, imaginaremos que Alice lleva a cabo una deliberación iterativa acerca de si debe mentir o decir la verdad. No obstante, nos gustaría efectuar una aclaración previa. Desde una perspectiva temporal, el autor parece equiparar la decisión de Alice con su acción, de lo cual podemos deducir que está concibiendo la decisión como una decisión proximal, tomada justamente antes de llevar a cabo la acción. Sin embargo,  $\delta'$  es una tesis relativa a decisiones distales, es decir, aquellas que se toman a medio o largo plazo, de manera que, para poder enfrentar nuestra tesis a la versión del argumento de *Mind* planteada por van Inwagen, nos tomaremos la licencia de suponer que la decisión de Alice es de tipo distal. Dando por buena esta modificación, que, a nuestro juicio, no afecta a la esencia ni a la potencia del ejemplo de van Inwagen, asumiremos que, si bien la reversión del tiempo propuesta por este tiene lugar a un instante situado un minuto antes de la decisión de Alice, esta no es ejecutada inmediatamente, sino cierto tiempo después. Por otro lado, y con el objetivo de situar nuestra argumentación, comenzaremos imaginando una situación en la que Dios da marcha atrás al mundo *hasta el instante en que Alice comienza a deliberar*, y no a un minuto antes de su decisión.

Imaginemos las dos primeras decisiones de Alice. Para comenzar con la primera, supongamos que, a partir de un instante  $t_0$ , la agente afronta el conflicto intencional entre las posibilidades alternativas “decir la verdad” y “mentir”. En aras de resolver el conflicto, nuestra imaginaria amiga comienza a deliberar, pero, dada la complejidad del conflicto y su falta de tiempo para resolverlo en un solo proceso reflexivo (iteración 1), tendrá que dejar el asunto aparcado hasta una mejor ocasión (iteración 2). Esta situación se repite en varias ocasiones de tal modo que, en un instante  $t_1$ , Alice toma la decisión de

decir la verdad tras, por ejemplo, 5 iteraciones. Ahora, y asumiendo las tres dimensiones del indeterminismo de la poda deliberativa, realizaremos algunas observaciones sobre la deliberación iterativa que Alice ha efectuado.

En primer lugar, diremos que, durante todo el intervalo temporal que ha mediado entre  $t_0$  y  $t_1$ , Alice desconocía cuál sería su propio proceder en relación a ser sincera o mentir. Esta incertidumbre interna ante su decisión futura ha servido como motivación a nuestra amiga para continuar deliberando hasta alcanzar dicha decisión. Estamos ante la *dimensión subjetiva* del indeterminismo en la poda deliberativa. En segundo lugar, diremos que la distribución sináptica del cerebro de Alice en el instante  $t_0$  (es decir,  $d_0$ ) y las leyes de la física que subyacen a los mecanismos celulares de cada neurona no han determinado la distribución sináptica que se presentaría tras las 5 iteraciones (es decir,  $d_5$ ). No obstante, y siguiendo la terminología empleada por Bishop unos párrafos más arriba, diremos que han delimitado “un espacio de posibilidades físicas” para la evolución de la distribución sináptica. Esta evolución ha estado decisivamente marcada por la influencia del contexto interno. Estamos aquí ante la *dimensión física* del indeterminismo en la poda deliberativa. Ahora bien, recordemos que, según proponíamos, dicha influencia se produce de dos maneras distintas, estableciendo con ellas dos dimensiones del indeterminismo físico: la normativa y la aleatoria. La *dimensión normativa* se refiere a la influencia decisiva del componente normativo, no sujeto a leyes naturales, y que en este caso ha consistido en el concepto que Alice tenía de sí misma, es decir, en su “yo”. Esta dimensión se ha presentado durante las 5 iteraciones deliberativas llevadas a cabo por Alice. La *dimensión aleatoria*, por su parte, se refiere a la influencia decisiva del cambio de contexto provocado por la interrupción de las constricciones normativas establecidas por el “yo” de nuestra protagonista. Este cambio, que se ha producido durante los 4 interciclos del proceso deliberativo, ha supuesto la aparición de efectos colectivos azarosos en la distribución en virtud de la actuación de la deriva sináptica. Finalmente, y en tercer lugar, diremos que la interacción entre las

dimensiones normativa y aleatoria de la dimensión física ha provocado la aparición de una *dimensión cíclica*, según la cual el número de iteraciones que Alice ha necesitado para resolver el conflicto intencional, que ha resultado ser 5, era incierto en el instante  $t_0$ .

A la vista de las afirmaciones anteriores, ¿puede afirmarse que Alice ha sido autora última de su primera decisión, es decir, que ha tenido control último sobre ella? En nuestra opinión, la respuesta es positiva. Es cierto que la deriva sináptica, interviniente en los interciclos, constituía un factor aleatorio y, por lo tanto, ajeno al control de Alice. Sin embargo, y a pesar de su injerencia estocástica, la deriva no puede decantar un conflicto intencional hacia una  $P$  concreta en tanto en cuanto es capaz de aumentar la variabilidad, no de reducirla. Así, Alice ha acabado por llegar a una decisión porque la  $P$  vencedora (“decir la verdad”) ha alcanzado una probabilidad de selección igual a 1 en virtud de su reiterada adecuación (hasta en 5 ocasiones, o iteraciones) al estándar normativo establecido por el “yo”. En este sentido, cabe decir que Alice, a través de su propio concepto de sí misma, ha sido autora de su decisión. También cabe sostener que el proceso estaba indeterminado: podía desarrollarse de varias formas, atendiendo a las condiciones impuestas por la naturaleza en  $t_0$ . Finalmente, puede decirse que Alice ha sido, de un modo cognitivo, responsable de sí misma y sus motivos porque su autoconcepto y las creencias evaluativas sobre sí misma, contenidas por él, se han formado en virtud de la sensibilidad y la coherencia de la agente hacia las consecuencias de sus acciones. Así, podemos concluir que Alice ha gozado de *control último* sobre su primera decisión.

Imaginemos ahora la segunda decisión de Alice. Dios revierte el tiempo hasta el instante  $t_0$  y respeta tanto el estado del mundo como las leyes de la naturaleza que existían antes de la primera deliberación. Desde el instante  $t_0$ , Alice delibera (iteración 1) para zanjar el conflicto entre “decir la verdad” y “mentir”, pero, debido a la complejidad del conflicto y a la falta de tiempo, deja el asunto aparcado hasta otra ocasión (iteración 2). La situación se repite de tal

manera que, en un instante  $t_2$ , Alice toma la decisión de mentir y lo hace, en este caso, tras 10 iteraciones. Nótese que Alice ha necesitado el doble de iteraciones que en la primera decisión, lo cual implica que, probablemente, el intervalo temporal que ha mediado entre  $t_0$  y  $t_2$  ha sido distinto al que anteriormente medió entre  $t_0$  y  $t_1$ . Esta diferencia temporal ha permitido que, a lo largo de esta segunda deliberación, el autoconcepto de Alice diverja del que albergaba durante la deliberación iterativa original. En los interciclos de la deliberación, distintos a los de entonces, la asimilación de las consecuencias de las acciones se ha producido de otro modo, desarrollando Alice otro “yo”. Tanto el autoconcepto como las creencias evaluativas que sobre sí misma tiene Alice, y que son albergadas por él, se han formado gracias a la sensibilidad y la coherencia de nuestra amiga hacia las consecuencias de sus acciones, haciéndola responsable de ambos. Por otro lado, es posible que dicho autoconcepto no se diferencie en exceso del anterior, pero quizá sí en algunos matices lo suficientemente relevantes como para afectar al estándar normativo que interviene en el conflicto intencional. Y esta divergencia normativa ha sido decisiva para que el resultado final haya cambiado. Ahora, y dando por válidas de nuevo las tres dimensiones del indeterminismo que hemos planteado antes, podemos realizar algunas afirmaciones sobre la deliberación iterativa que Alice ha concluido.

En primer lugar, durante el tiempo que ha mediado entre  $t_0$  y  $t_2$ , Alice desconocía cuál sería su propio proceder en relación a ser sincera o mentir, y esta incertidumbre la ha motivado para continuar deliberando hasta alcanzar la decisión. Esta es la *dimensión subjetiva* del indeterminismo presente en la deliberación de Alice. En segundo lugar, la distribución sináptica  $d_0$  y las leyes de la física no han determinado  $d_{10}$  (la distribución que se presentaría tras las 10 iteraciones), pero han delimitado un espacio de posibilidades físicas para la evolución de la distribución sináptica entre ambos extremos, evolución que ha estado decisivamente influenciada por el contexto interno. Esta es la *dimensión física*, que se manifiesta de dos maneras. Por un lado, la *dimensión normativa* se

refiere a la influencia decisiva del componente normativo, no sujeto a leyes naturales, y que, nuevamente, ha consistido en el autoconcepto de Alice. Esta dimensión se ha presentado durante las 10 iteraciones deliberativas. Por otro lado, la *dimensión aleatoria* se refiere a la influencia decisiva del cambio de contexto ocasionado por la interrupción de las constricciones normativas establecidas por el “yo” de Alice. Este cambio, que ha tenido lugar durante los 9 interciclos, ha supuesto la aparición de efectos colectivos azarosos en la distribución en virtud de la actuación de la deriva sináptica. Finalmente, y en tercer lugar, diremos que la interacción entre las dimensiones normativa y aleatoria ha provocado la aparición de la *dimensión cíclica*, por la que el número de iteraciones necesarias para que Alice resolviera el conflicto intencional, que ha resultado ser 10, era incierto en  $t_0$ .

¿Puede decirse que Alice ha sido, también ahora, la autora última de su decisión? De nuevo, pensamos que la respuesta es positiva, por las mismas razones que sostuvimos para la primera decisión. La deriva sináptica, aun siendo una influencia aleatoria, no decanta un conflicto intencional hacia una  $P$  específica porque no puede reducir la variabilidad. Alice ha llegado a una decisión porque la  $P$  vencedora (“mentir”) ha alcanzado una probabilidad de selección igual a 1 en virtud de su reiterada adecuación (durante 10 iteraciones) al estándar normativo establecido por el “yo”. De este modo, Alice ha sido autora de su decisión en virtud del concepto que alberga sobre sí misma. Además, el proceso estaba de nuevo indeterminado. Finalmente, cabe pensar que Alice ha sido cognitivamente responsable de sí misma y sus motivos porque su “yo” y las creencias evaluativas sobre sí misma, albergadas en este, se han formado gracias a la sensibilidad y la coherencia de la agente hacia las consecuencias de sus acciones. Así, Alice ha gozado de *control último* sobre su segunda decisión.

Según nuestro razonamiento, por tanto, la cuestión de qué decisión adopta Alice no es atribuible al azar, sino principalmente a la coherencia que existe entre el contenido de cada decisión y el “yo” de la agente a lo largo de las

deliberaciones iterativas correspondientes. A nuestro entender, esta afirmación sería válida cualquiera que sea el número de veces en que el tiempo dé marcha atrás, ya que, según postula la poda deliberativa, el factor aleatorio (la deriva sináptica) no puede nunca decantar un conflicto hacia un resultado específico. Además, la indeterminación en nuestro ejemplo no alude sólo al *contenido* de la decisión (“decir la verdad” o “mentir”), sino también al *momento* de tomarla (figura 5.8.). En este sentido, hubiéramos podido plantear también la posibilidad de que las dos decisiones de Alice hubieran guardado idéntico contenido (por ejemplo, “mentir” en ambos casos) pero hubieran sido tomadas tras un número diferente de iteraciones, y por lo tanto en momentos distintos. Por desgracia, el argumento de *Mind*, en sus diversas versiones, considera decisiones o acciones alternativas pero no *plazos alternativos*.<sup>30</sup> Sin embargo, semejante indeterminismo resulta clave para enfrentar la poda deliberativa a la concepción original del argumento propuesto por van Inwagen, según la cual Dios revierte el tiempo al estado del universo existente *un minuto antes de la primera decisión de Alice*. La interacción entre las dimensiones normativa y aleatoria, que da lugar a la dimensión cíclica del indeterminismo de la poda deliberativa, tiene lugar a lo largo de todo el proceso deliberativo, por lo que cabe afirmar que el número de iteraciones necesarias para que Alice resuelva el conflicto intencional es incierto *en cada instante* de dicho proceso, y no sólo en su inicio. No obstante, y según hemos sostenido, dicho conflicto sólo podrá llegar a su resolución en virtud del control de Alice, y no de la intervención de factores estocásticos.<sup>31</sup> En definitiva, Alice conservará siempre el control

---

<sup>30</sup> Así sucede, por ejemplo, en un caso desarrollado por Mele (1999, pp. 98-9). Este caso, que constituye otra de las versiones más conocidas del argumento, trata sobre dos gemelos habitan mundos idénticos en cuanto a su pasado y a las leyes de la naturaleza que los rigen.

<sup>31</sup> Obviamos aquí la nada despreciable posibilidad de que algún factor externo impida que Alice tome finalmente una decisión. Por ejemplo, nuestra protagonista podría morir antes de resolver el conflicto intencional, o podría encontrarse con que dicho conflicto se disuelve por sí mismo. Pero en casos como estos no se produce decisión alguna, por lo que carece de sentido preguntarse si Alice ha gozado de control último sobre ella.

último de su decisión: 1) independientemente del momento al que la deliberación regrese una y otra vez, y 2) a lo largo de la vía deliberativa correspondiente a cada reproducción. Consideramos, por tanto, que la tesis  $\delta'$  es inmune al argumento de *Mind*.

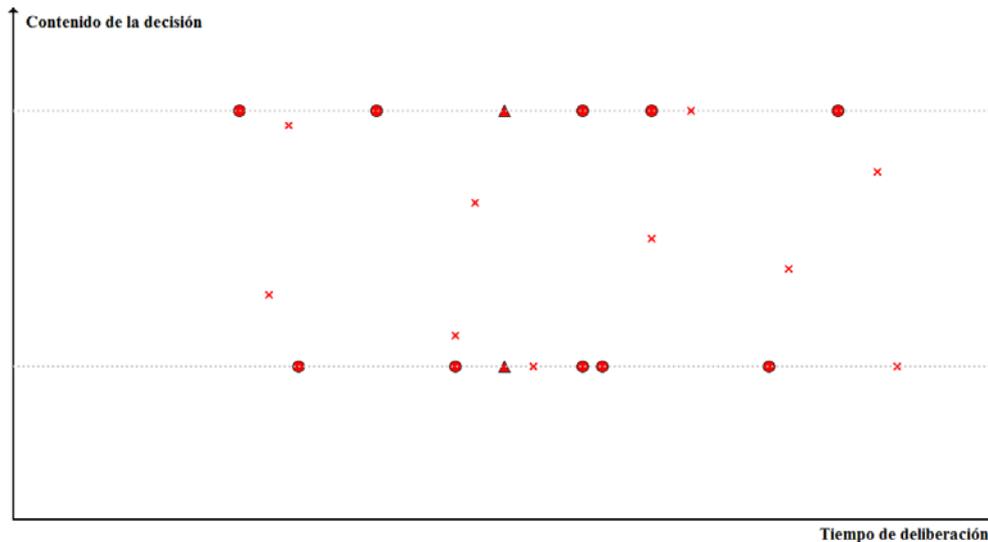


Figura 5.8. Nube de puntos en la que se combinan dos variables del indeterminismo en las decisiones de Alice. Cada punto representa una decisión ficticia. El ejemplo de van Inwagen no presta atención a la dimensión temporal, por lo que permite la posibilidad de pensar que las decisiones de Alice (“mentir” o “decir la verdad”) sólo pueden repartirse entre dos ubicaciones (puntos triangulares). Según la poda deliberativa, en cambio, las dos alternativas de contenido pueden presentarse en diferentes instantes, convirtiendo a cada decisión en un evento único o, cuanto menos, difícilmente repetible (puntos circulares). Otra posibilidad, no mencionada en el texto, consiste en entender que las decisiones “mentir” y “decir la verdad” se sitúan en los extremos de un intervalo continuo en el que caben soluciones más o menos intermedias (*e.g.*, “decir una verdad parcial”) repartidas en diversas ubicaciones temporales (puntos con forma de aspa).<sup>32</sup>

<sup>32</sup> Hay espacio para la variabilidad incluso dentro de cada extremo del intervalo, pues cabe la posibilidad de que Alice mienta de diversos modos e, igualmente, de que diga la verdad de maneras distintas (cf. Moya 2006b, p. 196; véase apartado 4.3.2.).

No querríamos concluir este capítulo sin reconocer que varios de los planteamientos de Moya nos han servido de especial inspiración a la hora de desarrollar nuestra doble aproximación: las tesis  $\delta$  y  $\delta'$ . No en vano, y como ya sostuvimos en la sección 5.3., el concepto de poda semántica es altamente compatible con la propuesta que este autor defiende para la conciliación entre materialismo y causalidad intencional. Además, cabe señalar nuevamente que hemos tomado prestada de Moya la noción general de causalidad a la que se refiere nuestra tesis preliminar,  $\alpha$ , de la que dependen en última instancia tanto  $\delta$  como  $\delta'$ . Finalmente, cabe recordar que el rol normativo desempeñado por el autoconcepto o “yo” del agente en la poda deliberativa guarda similitud con respecto al que ejercen las creencias evaluativas en la propuesta libertarista de nuestro autor.

## Conclusiones

«- ¡Creo que está perdido!  
- ¿Ah, sí?  
- ¿A dónde se dirige?  
- Precisamente ahora iba a decidirlo.»

*Náufrago* (Robert Zemeckis, 2000).<sup>1</sup>

En este trabajo hemos abordado el problema de la relación entre el determinismo y el libre albedrío y la responsabilidad moral empleando un enfoque basado en el conocimiento científico contemporáneo y, en especial, en el propio de la neurociencia. Partiendo de dicho enfoque, hemos efectuado una propuesta libertarista que puede ser clasificada dentro del materialismo no reductivo. A lo largo de nuestra investigación, siempre hemos procedido con el convencimiento de que resulta imprescindible que la mentalidad *lineal, reduccionista y mecanicista* se deje complementar por las explicaciones *no lineales, holísticas y sistémicas* (cf. Murphy 2011, p. 212). En este sentido, estamos plenamente de acuerdo con Bishop (2008, p. 245) cuando afirma:

[L]as explicaciones filosóficas típicas de la causalidad [...] fueron desarrolladas dentro del contexto de una “visión lineal del mundo” fuertemente influida por un enfoque casi exclusivo en los modelos lineales de la mecánica newtoniana. En el contexto de los sistemas no lineales [...] hay muchos canales adicionales de interacción no concebidos en tal marco lineal (e.g., sistemas abiertos intercambiando materia y energía con sus entornos (una forma de situacionalidad), bucles de retroalimentación no lineales,

---

<sup>1</sup> El guión de esta película es obra de William Broyles Jr. El fragmento transcrito corresponde a la versión doblada al castellano.

interacciones internivel, autoorganización, relaciones todo-parte, efectos de memoria). Tratar con estas formas adicionales de interacción parece difícil sin modificar o extender las explicaciones causales típicas.

Si echamos un vistazo al debate puramente filosófico sobre la relación entre el determinismo y el libre albedrío (y/o la responsabilidad moral), nos daremos cuenta de que podría alargarse mucho tiempo aún. En el capítulo 1 hemos situado el marco general de dicho debate y hemos sugerido una organización sistemática de las principales posturas existentes. Las ideas fundamentales que podemos extraer son las siguientes:

[I] Los seres humanos poseemos un concepto *a priori*, basado en nuestras intuiciones, sobre una capacidad que conocemos comúnmente como “libertad”, “libre voluntad” o “libre albedrío”. Desde un punto de vista preliminar, podemos atribuirle a dicho concepto dos características fundamentales: (1) posibilidades alternativas de decisión y/o acción, y (2) control último sobre nuestras decisiones y/o acciones.

[II] En la literatura acerca de la libertad y la responsabilidad moral, a menudo se han tratado ambos temas conjuntamente, y en muchas ocasiones se ha considerado al libre albedrío como condición indispensable para la responsabilidad moral. No obstante, no existe un consenso generalizado acerca de esta relación.

[III] Tanto el compatibilismo como el incompatibilismo se sustentan en los cimientos de una verdad lógica, la premisa ( $\alpha$ ): el determinismo es cierto o no lo es.

[IV] La incompatibilidad del determinismo con la libertad puede ser presentada como conclusión de dos premisas: (*a*) el determinismo impide la existencia de posibilidades alternativas de decisión y/o acción, y (*b*) las posibilidades alternativas son necesarias para decidir y/o actuar libremente. Alternativamente, puede ser presentada como conclusión de otras dos premisas: (*c*) el determinismo impide el control último sobre nuestras

decisiones y/o acciones, y *(d)* el control último es necesario para decidir y/o actuar libremente.

[V] Las mayoría de las versiones del compatibilismo se basan en la negación de alguna de las cuatro premisas *(a)*, *(b)*, *(c)* y *(d)*, sin negar por ello la existencia del libre albedrío. La teoría de Peter Strawson acerca de las actitudes reactivas no puede situarse, en cambio, en el marco de dichas premisas.

[VI] El incompatibilismo en su versión libertarista acepta *(a)*, *(b)*, *(c)* y *(d)*, y sostiene que el indeterminismo es compatible con la libertad y necesario para que esta se dé. Esto conduce a los libertaristas a la obligación de ofrecer una explicación convincente a la denominada “Cuestión de Inteligibilidad” (¿es inteligible un libre albedrío indeterminista?).

[VII] Las versiones no libertaristas del incompatibilismo son críticas con el compatibilismo y con el libertarismo a partes iguales: niegan la compatibilidad entre determinismo y libertad, pero también la inteligibilidad del libre albedrío concebido por los libertaristas.

[VIII] Existen posturas acerca de la libertad y la responsabilidad moral que se desmarcan de la división tradicional entre compatibilismo e incompatibilismo y para las cuales las premisas *(a)*, *(b)*, *(c)* y *(d)* no son de aplicación. Se trata del autonomismo agnóstico de Mele (no hace falta decidirse por el compatibilismo o el incompatibilismo para creer que existen personas autónomas), el subjetivismo de Double (es subjetivo, por estar basado en actitudes morales, atribuir la capacidad de libre albedrío a una persona) y el revisionismo de Vargas (lo que tendemos a creer acerca de la libertad y/o la responsabilidad moral es diferente a lo que deberíamos creer).

En definitiva, cada una de las distintas posiciones cree captar nuestras intuiciones acerca de qué es la libertad (y/o la responsabilidad moral), y considera que ofrece una explicación adecuada a dichas intuiciones. En este sentido, resulta difícil imaginar una solución cercana al debate filosófico alrededor de la relación entre determinismo y libertad. En cualquier caso, el problema necesitaba, a nuestro juicio, de una organización sistemática que

serviera para enmarcar el resto del trabajo. Ese ha sido nuestro objetivo en el capítulo 1.

A la vista del embarrado jardín conceptual que se presenta ante nuestros ojos, parece razonable acudir a la ciencia y averiguar si puede ayudarnos a encontrar respuestas. Aunque el trabajo filosófico es del todo imprescindible, no es menos cierto que necesita de la ciencia para no estancarse. En los capítulos 2 y 3 hemos realizado un estudio crítico sobre la relación que existe entre el conocimiento científico actual y el problema del libre albedrío. El análisis se ha llevado a cabo desde puntos de vista diferentes representados por cinco cuestiones fundamentales propuestas por Robert Kane. Recordémoslas:

[A] La Cuestión de Compatibilidad: ¿Es el libre albedrío compatible con el determinismo?

[B] La Cuestión de Significatividad: ¿Por qué queremos, o deberíamos querer, poseer un libre albedrío que es incompatible con el determinismo? ¿Es este un tipo de libertad “que valga la pena querer” (por utilizar la útil frase de Dennett)? Y si es así, ¿por qué?

[C] La Cuestión de Inteligibilidad: ¿Podemos darle sentido a una libertad o libre albedrío que es incompatible con el determinismo? ¿Es dicha libertad coherente o inteligible? ¿O es, como reivindicán muchos críticos, esencialmente misteriosa y terminantemente oscura?

[D] La Cuestión de Existencia: ¿Dicha libertad existe en realidad en el orden natural? Y si es así, ¿dónde?

[E] La Cuestión Determinista: ¿Es el determinismo cierto?

Hemos comenzado fijando nuestra atención sobre el nivel mecánico-cuántico. Del estudio al respecto, que hemos efectuado en el capítulo 2, nos gustaría extraer tres conclusiones generales:

[I] Es posible que las indeterminaciones cuánticas proporcionen una fuente para el surgimiento de posibilidades alternativas entre las cuales un agente pueda realizar una elección. Esta posibilidad, por sí sola, otorga relevancia a la mecánica cuántica en el debate sobre el libre albedrío, independientemente de la relación que exista entre la condición de control último y los fenómenos cuánticos.

[II] No sabemos con certeza si el indeterminismo cuántico tiene origen ontológico o epistemológico. Encontrar una respuesta concluyente a la Cuestión Determinista resulta fundamental para contextualizar las discusiones entre compatibilistas e incompatibilistas, y la mecánica cuántica todavía tiene mucho que decir sobre dicha cuestión.

[III] Suponiendo un indeterminismo ontológico de la mecánica cuántica, los libertaristas podrían integrarlo en sus argumentos, aunque se encontrarían con el difícil obstáculo de tener que explicar la condición de control último para un tipo de libertad con origen intrínsecamente azaroso. La hipótesis de Hameroff y Penrose tendría posibilidades de ayudar a sortear dicho obstáculo, aunque a día de hoy no goza de apoyo empírico. Otras hipótesis aquí expuestas muestran demasiados inconvenientes como para ser de ayuda al libertarismo a día de hoy. En cualquier caso, cualquier respaldo cuántico a las hipótesis libertaristas debería encaminarse hacia un sentido *descendente* de la causalidad, es decir, del nivel mental al nivel físico. De esta manera, los fenómenos cuánticos constituirían un correlato de la libertad del agente... pero no como su *fuentes* sino como su *consecuencia*. Sólo así podría eludirse el argumento de *Mind*, que proclama la ausencia de control último en una voluntad surgida del azar.

En el capítulo 3, hemos continuado con nuestro estudio crítico fijándonos en el nivel neural. Hemos podido constatar que las técnicas de la neurociencia, entre las que destacan los estudios de lesiones, de neuroimagen funcional y de neuroelectrofisiología, han permitido ahondar en los mecanismos cerebrales relacionados con la volición. Sin embargo, habiéndonos aproximado al debate sobre el libre albedrío con la inspiración de la neurofilosofía mínima, sin posicionamiento metafísico previo acerca del problema mente-cuerpo, la conclusión general que podemos extraer es que la neurociencia actual no nos permite alcanzar una solución a dicho debate (cf. Mele 2009, p. 161; cf. Mele 2011, p. 513-14), y que por ello, sin menoscabo de la necesaria ayuda de la reflexión filosófica, debe ahondar en sus avances,

especialmente en lo relativo a *la neuroplasticidad, las decisiones distales y la deliberación*. Esta conclusión parte principalmente de tres argumentos:

[I] Algunas estructuras nerviosas, entre las que el córtex prefrontal es especialmente importante, parecen ser responsables de la función volitiva. También hay que considerar la estrecha conexión de la voluntad con la emoción. Si a esto le unimos el hecho de que el trabajo de Libet y Wegner apunta hacia un gran peso de los procesos inconscientes en nuestros actos, y además consideramos el influjo causal descendente del entorno sociocultural sobre la voluntad y las vías neurales del individuo, una primera impresión podría apuntar a que nuestra voluntad sería explicable mediante teorías deterministas. Sin embargo, hay razones para no abandonar el indeterminismo:

(a) Resulta problemático hallar un origen causal último de nuestra voluntad a nivel anatómico y, además, la actividad de las áreas nerviosas implicadas es condición necesaria pero no obligatoriamente suficiente para el proceso volitivo.

(b) Las posturas de Libet y Wegner han recibido diversas críticas (si bien el primero no descarta el indeterminismo).

(c) La causalidad descendente, cuya existencia es motivo de discusión, no es incompatible con mecanismos indeterministas.

[II] Libet deja la puerta abierta a la posibilidad de que ejerzamos nuestra libertad mediante el veto consciente de algunas opciones de acción frente a otras. Los libertaristas podrían, en principio, utilizar dicha posibilidad para apoyar sus teorías, pero se encontrarían con las siguientes dificultades:

(a) Es necesario explicar cómo es posible que se pueda seleccionar conscientemente entre opciones de naturaleza inconsciente.

(b) Cabe la posibilidad de que el mecanismo de veto tenga lugar sin que exista un control último.

(c) Hay que aclarar cómo podría el indeterminismo originar tanto mecanismos inconscientes (surgimiento de opciones de acción) como conscientes (proceso de veto).

[III] Aunque tomáramos la obra de Libet y de Wegner como un apoyo para defender el determinismo de nuestras decisiones y nuestras acciones, hay razones para sostener que la neurociencia no nos permite descartar la existencia de un libre albedrío incompatible con el determinismo, a pesar de las dificultades para aceptarla. A saber:

(a) La neuroplasticidad que nuestro cerebro es capaz de mostrar en algunas ocasiones abre la posibilidad de que mecanismos indeterministas aún no descubiertos puedan moldear nuestro cerebro y modificar nuestras capacidades al ejercer nuestra libertad (si bien estos mecanismos necesitarían de una descripción complementaria que explicara coherentemente la causalidad mental, de tipo descendente, y el control último por parte del agente).

(b) No podemos dejar a un lado el riesgo que supone explicar decisiones relevantes de tipo distal de la misma forma en que explicamos decisiones proximales irrelevantes, como las que aparecen en los experimentos de Libet (Mele 2011, p. 507).

(c) No debemos olvidar la importancia de la deliberación para nuestras decisiones y nuestros actos, tratada secundariamente por Libet y por Wegner en beneficio de la espontaneidad.

En el capítulo 4, hemos analizado las opciones que tiene el libertarismo de hallar una teoría compatible con el conocimiento científico actual. Hemos sugerido que semejante teoría debería satisfacer tres condiciones fundamentales. La *causalidad mental*, que era la primera de ellas, parece enfrentarse a un conflicto con cinco tesis: (a) el anomalismo en el marco de lo mental, (b) el externismo semántico e intencional, (c) el funcionamiento estrictamente sintáctico del cerebro, concebido como el órgano responsable del comportamiento intencional, (d) el cierre causal de la física, y (e) la concepción jerárquica de la realidad que parte del nivel físico como metafísicamente fundamental. En consecuencia, para poder cumplir con la condición de causalidad mental se hace imprescindible que, en relación a cada una de las

mencionadas tesis, o bien se rebata la tesis, o bien se niegue razonadamente que exista un conflicto real entre la tesis y la causalidad mental. Además, cualquier análisis de la causalidad mental debe prestar su justa atención al vínculo entre razones y acción intencional y al papel de los estados fenomenológicos.

La segunda condición no era otra que la *correlación inteligible entre la causalidad mental y un sustrato material*. Para satisfacerla es necesario integrar explicaciones mentales y materialistas en un único marco teórico. A este respecto, hemos presentado los intentos de Dretske, Van Gulick y Moya. También nos hemos preguntado por una descripción adecuada de dos elementos fundamentales:

- a) un vínculo metafísico entre propiedades de niveles jerárquicos distintos que conlleve una relación causal descendente entre ellas.
- b) un sustrato material a través del cual las propiedades de nivel superior operen sobre las propiedades de niveles inferiores.

En cuanto al primer requisito, hemos argumentado que la relación adecuada es la emergencia contextual, en la que la descripción de propiedades en un nivel jerárquico particular es necesaria pero insuficiente para llegar a la descripción en un nivel superior, siendo imprescindible acudir al contexto para dicha tarea. En cuanto al sustrato, hemos sostenido que el mejor candidato son los sistemas complejos, que presentan, entre otras características interesantes, emergencia contextual y un tipo de causalidad descendente que, en función del tipo de contexto actuante, podía ser vista como:

- 1) causalidad formal, al estilo de Van Gulick, cuando actúa un contexto interno.
- 2) causalidad estructurante, al estilo dretskeano, cuando actúa un contexto normativo.

Precisando más, hemos sugerido que, concibiendo sistema nervioso y mente como dos niveles jerárquicos consecutivos dentro de un único sistema complejo, el organismo humano, sucede que *determinadas constricciones normativas*

*de naturaleza semántica actuando sobre el nivel mental provocan que se seleccionen y refuerzen rutas neurales y neuromusculares conducentes a la ejecución de ciertas acciones.* Hemos destacado que esta concepción de la causalidad mental se adecua a los intentos de Dretske y de Moya.

La tercera condición consistía en una *indeterminación ontológica que permita el control último*. Para ilustrar la estructura de esta condición hemos expuesto dos vías alternativas de concebir el origen del indeterminismo y su relación con el control último. La primera vía, planteada por Kane, postula un origen interno del indeterminismo, y sigue una perspectiva individualista y centrada en la voluntad. La segunda vía, propuesta por Moya, se basa en un enfoque social y cognitivo en el que el indeterminismo surge del ámbito normativo, externo al agente. Parece que esta segunda vía resulta más fácilmente inteligible que la primera.

El análisis de las tres condiciones mencionadas nos ha servido como cimiento para, ya en el capítulo 5, elaborar una propuesta propia de carácter libertarista y compatible con los postulados materialistas en los que se basa la ciencia contemporánea. Al respecto, hemos procedido en dos pasos: la poda semántica (tesis  $\delta$ ) y la poda deliberativa (tesis  $\delta'$ ). Con la primera tesis hemos propuesto un posible modelo de actuación del contenido y el significado sobre el mundo físico. Ello nos ha permitido sentar las bases para la segunda tesis, con la que hemos planteado un modelo paralelo vinculado a la toma interna de decisiones y con carácter marcadamente indeterminista. Dicho modelo, que constituye el resultado final de nuestra tesis doctoral, postula que *determinadas constricciones normativas de naturaleza semántica y con estructura de bucle extraño actuando sobre el nivel mental provocan que se dé un proceso de poda sináptica en el que se seleccionan y refuerzan rutas neurales y neuromusculares conducentes a la ejecución de ciertas acciones.*

Aunque nuestro doble modelo (poda semántica y poda deliberativa) se basa específicamente en la poda sináptica, no descartamos formas de actuación *complementarias* sostenidas sobre mecanismos neurobiológicos distintos, en

especial sobre aquellos propios de la neuroplasticidad. También cabría valorar la posibilidad de que fenómenos de neuroplasticidad “blanda”, en los que meramente se debilitan ciertas conexiones., antecedan a una fase de poda sináptica, mucho más agresiva en tanto en cuanto en ella se erradican itinerarios sinápticos completos. En cualquier caso, somos conscientes de las obvias limitaciones impuestas por el carácter especulativo de nuestra aproximación, por lo que queda lejos de nuestra intención pretender que esta resuelva el problema de la explicación de lo mental e intencional en términos libertaristas y materialistas. Pensamos, no obstante, que es en parte susceptible de ser verificada empíricamente. La velocidad a la que se vienen desarrollando los métodos de mapeo de las estructuras y conexiones del cerebro, y del sistema nervioso en general, permite ser optimistas en este sentido.

En el caso de la tesis  $\delta$  (poda semántica), la verificación requeriría, probablemente, el empleo de técnicas *in vivo* que permitan realizar en un grupo experimental el seguimiento de la evolución de las rutas sinápticas situadas en áreas o subáreas nerviosas específicas durante un proceso de aprendizaje de destrezas sometidas a reglas de significado, así como el contraste con respecto a las mismas áreas o subáreas en un grupo de control. Consideramos que sería una técnica más adecuada que la medición de la densidad de sinapsis, la cual no permite detectar posibles procesos de poda sináptica poco relevantes cuantitativamente, que permanecerían ocultos dentro del balance total entre la creación y la destrucción de sinapsis. Sin embargo, tales procesos podrían ser significativos desde el punto de vista cualitativo. Este problema experimental resulta particularmente importante en el caso de los individuos en plena edad adulta, dado que nuestra propuesta exige que la poda sináptica se perpetúe más allá de la fase de desarrollo nervioso, algo sobre lo que se ha especulado entre los neurobiólogos pero que no ha podido ser corroborado hasta el momento.

En cuanto a la tesis  $\delta'$  (poda deliberativa), cabe reconocer que la verificación se nos antoja más complicada. La propuesta conlleva un enfoque holístico según el cual se produce una redistribución de botones sinápticos a

gran escala en el cerebro, y un estudio a nivel cerebral resulte inabarcable en la práctica. Una alternativa podría consistir en la verificación indirecta a través de indicios favorables a la tesis. Por ejemplo, encontramos plausible efectuar un estudio similar al que hemos planteado para la poda semántica. Dicho estudio consistiría en un seguimiento *in vivo* de la evolución de la distribución sináptica en áreas o subáreas del córtex prefrontal, del cual se sabe que está fuertemente vinculado a la deliberación, el pensamiento complejo y la toma de decisiones. El seguimiento requeriría el contraste entre un grupo de control y un grupo experimental compuesto, por ejemplo, por individuos que, dada su profesión, se ven empujados a deliberar con frecuencia y regularidad elevadas (*e.g.*: jueces, filósofos, detectives, etc.). En este sentido, unos resultados que reflejaran una redistribución sináptica significativamente más vertiginosa en el grupo experimental que en el grupo de control constituiría un indicio favorable a la poda deliberativa, en tanto en cuanto demostrarían que la deliberación es un factor promotor de la poda sináptica.

Dejando a un lado la posibilidad de su verificación empírica con seres humanos, se nos ocurren diferentes vías de profundización adicionales en relación a los modelos de poda semántica y poda deliberativa. Con las cuestiones que a continuación planteamos pretendemos, de manera breve, sugerir posibles líneas de investigación en el futuro:

[I] *Naturaleza de la mente y problema mente-cuerpo.* ¿Conviven las podas semántica y deliberativa con otras modalidades de causalidad mental? ¿Son compatibles con diferentes teorías acerca del problema mente-cuerpo? ¿Es posible, incluso, relacionar lo fenomenológico y lo intencional con la ayuda de nuestro doble modelo? Así, por ejemplo: ¿surgen ciertos eventos o estados intencionales a partir de eventos o estados fenomenológicos que quedan “atrapados” en bucles extraños?

[II] *Inteligencias animal y artificial.* ¿Podemos encontrar procesos de poda semántica en especies distintas a la humana (especialmente dentro de los primates o los cetáceos)? Es más, ¿existen antecedentes primitivos de la poda

deliberativa dentro del reino animal? ¿Es posible recrear informáticamente el doble modelo que proponemos? Y en caso afirmativo, ¿podría hallarse algún tipo de aplicación práctica en la investigación y el desarrollo de la inteligencia artificial?

[III] *Filosofía moral.* ¿Permite la poda deliberativa establecer una gradualidad en cuanto a la libertad de decisión y/o acción de un agente volitivo? Y si así fuera, ¿habría posibles consecuencias legales? ¿Podrían establecerse circunstancias agravantes o atenuantes para un delito atendiendo al “grado de libertad” de quien lo comete? ¿Sería moralmente aceptable establecer diferencias entre las personas por razón de su “grado de libertad”?

[IV] *Filosofía de la biología.* ¿Cabe relacionar nuestro doble modelo con los postulados defendidos por la teoría de la evolución? ¿Juegan las podas semántica y deliberativa algún papel evolutivo? ¿Ejercen algún tipo de presión selectiva dentro de una población humana? ¿Ha evolucionado la poda deliberativa a partir de la semántica?

Concluye aquí esta tesis doctoral. Con nuestro resultado final de investigación, la poda deliberativa, esperamos haber ofrecido una tesis libertarista que, cuando menos, resulte razonablemente inteligible a la vez que compatible con la neurociencia moderna. También esperamos haber puesto el suficiente énfasis en la neuroplasticidad, las decisiones distales y la deliberación, elementos que, según sosteníamos al final del capítulo 3, deberían ser estudiados con mayor profundidad por la neurociencia actual. En relación al tercero de dichos elementos, confiamos en que el lector haya encontrado realista el concepto de deliberación iterativa. Pensamos que, siempre que un conflicto intencional sea lo bastante complejo, las deliberaciones de esta clase podrían producirse con relativa frecuencia, más aún si pensamos en el ritmo al que gira nuestro mundo presente, privándonos a menudo de períodos de tiempo suficientemente largos para, entre otras cosas, poder deliberar sosegadamente. En este sentido, pensamos que la deliberación iterativa, que conlleva interrupciones y

reanudaciones del proceso reflexivo, encaja con nuestro estilo de vida contemporáneo.



## Bibliografia

- Anscombe, G.E.M. 1975. "Causality and Determination". En *Causation and Conditionals*, ed. E. Sosa, 63-81. Oxford: Oxford University Press.
- . 1963. *Intention (2nd ed.)*. Cambridge, Massachusetts: Harvard University Press.
- . 1957. *Intention*. Oxford: Blackwell.
- Antoniou, I.E., y I. Prigogine. 1993. "Intrinsic Irreversibility and Integrability of Dynamics". *Physica A* 192: 443-64.
- Aspect, A., J. Dalibard, y G. Roger. 1982. "Experimental Test of Bell's Inequalities Using Time-Varying Analyzers". *Physical Review Letters* 49: 1804-7.
- Atmanspacher, H. 2012. "Identifying Mental States from Neural States under Mental Constraints". *Interface Focus* 2: 74-81.
- Atmanspacher, H., y P. beim Graben. 2007. "Contextual Emergence of Mental States from Neurodynamics". *Chaos and Complexity Letters* 2: 151-68.
- Audi, R. 1993. "Mental Causation: Sustaining and Dynamic". En *Mental Causation*, eds. J. Heil y A. Mele, 53-74. Oxford: Clarendon Press.
- Bacciagaluppi, G., y A. Valentini. 2009. *Quantum Theory at the Crossroads: Reconsidering the 1927 Solway Conference*. Cambridge: Cambridge University Press.
- Balaguer, M. 2010. *Free Will as an Open Scientific Problem*. Cambridge, Massachusetts: MIT Press.
- Barrow, J.D., y F.J. Tipler. 1988. *The Anthropic Cosmological Principle*. Oxford: Oxford University Press.

- Beck, F., y J.C. Eccles. 1992. "Quantum Aspects of Brain Activity and the Role of Consciousness". *Proceedings of the National Academy of Sciences of the USA* 89: 11357-61.
- Bell, J.S. 1964. "On the Einstein Podolsky Rosen Paradox". *Physics* 1: 195-200.
- Bertrán, M. 1993. "El influjo de Hofstadter sobre una versión novedosa del compatibilismo". *Enrabonar* 20: 7-16.
- Bickle, J. 2013. "Multiple Realizability". *The Stanford Encyclopedia of Philosophy*, <<http://plato.stanford.edu/archives/spr2013/entries/multiple-realizability/>> (fecha de consulta: 02/11/2015).
- Bishop, R.C. 2012. "Excluding the Causal Exclusion Argument Against Non-Reductive Physicalism". *Journal of Consciousness Studies* 19: 57-74.
- . 2011a. "Chaos, Indeterminism, and Free Will". En *The Oxford Handbook of Free Will (2nd ed.)*, ed. R. Kane, 84-100. Nueva York: Oxford University Press.
- . 2011b. "Metaphysical and Epistemological Issues in Complex Systems". En *Handbook of the Philosophy of Science, Volume 10: Philosophy of Complex Systems*, ed. C. Hooker, 105-36. Amsterdam: North Holland.
- . 2008. "Downward Causation in Fluid Convection". *Synthese* 160: 229-48.
- . 2004. "Nonequilibrium Statistical Mechanics Brussels-Austin Style". *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 35: 1-30.
- . 2002. "Chaos, Indeterminism, and Free Will". En *The Oxford Handbook of Free Will*, ed. R. Kane, 111-24. Nueva York: Oxford University Press.
- Bishop, R.C., y H. Atmanspacher. 2011. "The Causal Closure of Physics and Free Will". En *The Oxford Handbook of Free Will (2nd ed.)*, ed. R. Kane, 101-11. Nueva York: Oxford University Press.
- . 2006. "Contextual Emergence in the Description of Properties". *Foundations of Physics* 36: 1753-77.

- Bohm, D. 1952a. "A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables I". *Physical Review* 85: 166-79.
- . 1952b. "A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables II". *Physical Review* 85: 180-93.
- Bohm, D., y B.J. Hiley. 1993. *The Undivided Universe: An Ontological Interpretation of Quantum Mechanics*. Abingdon, Reino Unido: Routledge.
- Bohr, N. 1913a. "On the Constitution of Atoms and Molecules". *Philosophical Magazine Series 6* 26: 1-25.
- . 1913b. "On the Constitution of Atoms and Molecules. Part II.- Systems Containing Only a Single Nucleus". *Philosophical Magazine Series 6* 26: 476-502.
- . 1913c. "On the Constitution of Atoms and Molecules. Part III.- Systems Containing Several Nuclei". *Philosophical Magazine Series 6* 26: 857-75.
- Boleyn-Fitzgerald, M. 2010. *Pictures of The Mind: What the New Neuroscience Tells Us about Who We Are*. Upper Saddle River, Nueva Jersey: FT Press.
- Born, M. 1926. "Zür Quantenmechanik der Stossvorgänge". *Zeitschrift Für Physik A: Hadrons and Nuclei* 37: 863-7.
- Brassard, G., y P. Raymond-Robichaud. 2013. "Can Free Will Emerge from Determinism in Quantum Theory?". En *Is Science Compatible with Free Will?: Exploring Free Will and Consciousness in the Light of Quantum Physics and Neuroscience*, eds. A. Suarez y P. Adams, 41-61. Nueva York: Springer.
- Bub, J. 1997. *Interpreting the Quantum World*. Cambridge: Cambridge University Press.
- Bunge, M. 1980. *The Mind-Body Problem: A Psychobiological Approach*. Oxford: Pergamon Press.
- Cabello, A. 2007. "Reducción de la comunicación mediante entrelazamiento: Del "descubrimiento más profundo de la ciencia" a la "pseudo-telepatía cuántica"". *Revista Española de Física* 21: 58-60.

- Campbell, D.T. 1974. “‘Downward Causation’ in Hierarchically Organised Biological Systems”. En *Studies in the Philosophy of Biology: Reduction and Related Problems*, eds. F.J. Ayala y T. Dobzhansky, 179-86. Berkeley y Los Ángeles: University of California Press.
- Cavalcanti, S., y F. Fontanazzi. 1999. “Deterministic Model of Ion Channel Flipping with Fractal Scaling of Kinetic Rates”. *Annals of Biomedical Engineering* 27: 682-95.
- Conway, J., y S. Kochen. 2006. “The Free Will Theorem”. *Foundations of Physics* 36: 1441-73.
- . 2009. “The Strong Free Will Theorem”. *Notices of the American Mathematical Society* 56: 226-32.
- Crutchfield, J.P. 1994. “Observing Complexity and the Complexity of Observation”. En *Inside versus Outside: Endo- and Exo-Concepts of Observation and Knowledge in Physics, Philosophy and Cognitive Science*, eds. H. Atmanspacher y G.J. Dalenoort, 235-72. Berlín: Springer.
- Cuesta, J.A. 2006. “La entropía como creadora de orden”. *Revista Española de Física* 20: 13-19.
- Chow, C.C., y J.A. White. 1996. “Spontaneous Action Potentials Due to Channel Fluctuations”. *Biophysical Journal* 71: 3013-21.
- Churchland, P.S. 1986. *Neurophilosophy: Toward a Unified Science of the Mind/ Brain*. Cambridge, Massachusetts: MIT Press.
- da Costa, N.C.A., y C. de Ronde. 2013. “The Paraconsistent Logic of Quantum Superpositions”. *Foundations of Physics* 43: 845-58.
- Damasio, A.R. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain*. Nueva York: Avon Books.
- Damasio, A.R., y G.W. van Hoesen. 1983. “Emotional Disturbances Associated with Focal Lesions of the Limbic Frontal Lobe”. En *Neuropsychology of Human Emotion*, eds. K.M. Heilman y P. Satz, 87-103. Nueva York: Guilford.

- Damasio, H., T. Grabowski, R. Frank, A.M. Galaburda, y A.R. Damasio. 1994. "The Return of Phineas Gage: Clues about the Brain from the Skull of a Famous Patient". *Science* 264: 1102-5.
- Danto, A.C. 1973. *Analytical Philosophy of Action*. Cambridge: Cambridge University Press.
- Davidson, D. 2001. *Essays on Actions and Events (2nd ed.)*. Oxford: Oxford University Press.
- . 1993. "Thinking Causes". En *Mental Causation*, eds. J. Heil y A. Mele, 3-17. Oxford: Clarendon Press.
- . 1963. "Actions, Reasons, and Causes". *The Journal of Philosophy* 60: 685-700.
- Davies, P.C.W. 2006. "The Physics of Downward Causation". En *The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion*, eds. P. Clayton y P.C.W. Davies, 35-52. Oxford: Oxford University Press.
- de Broglie, L. 1925. "Recherches sur la théorie des quanta". *Annales de Physique* 3: 22-128.
- de Ronde, C. 2012. "La noción de potencialidad ontológica en la interpretación modal de la mecánica cuántica". *Scientiae Studia* 10: 137-64.
- deCharms, R.C., F. Maeda, G.H. Glover, D. Ludlow, J.M. Pauly, D. Soneji, J.D.E. Gabrieli, y S.C. Mackey. 2005. "Control over Brain Activation and Pain Learned by Using Real-Time Functional MRI". *Proceedings of the National Academy of Sciences of the USA* 102: 18626-31.
- Dennett, D.C. 1991. *Consciousness Explained*. Londres: Penguin.
- . 1984. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, Massachusetts: MIT Press.
- Deutsch, D. 1998. *The Fabric of Reality*. Londres: Penguin.
- Diósi, L. 1988. "Quantum Stochastic Processes as Models for State Vector Reduction". *Journal of Physics A* 21: 2885-98.
- Dolan, R.J., C.J. Bench, P.F. Liddle, K.J. Friston, C.D. Frith, P.M. Grasby, y R.S.J. Frackowiak. 1993. "Dorsolateral Prefrontal Cortex Dysfunction in

- the Major Psychoses; Symptom or Disease Specificity?" *Journal of Neurology, Neurosurgery, and Psychiatry* 56: 1290-94.
- Double, R. 2002. "Metaethics, Metaphilosophy, and Free Will Subjectivism". En *The Oxford Handbook of Free Will*, ed. R. Kane, 506-28. Nueva York: Oxford University Press.
- Dretske, F. 2009. "What Must Actions Be for Reasons to Explain Them?". En *New Essays on the Explanation of Action*, ed. C. Sandis, 13-21. Basingstoke, Reino Unido: Palgrave Macmillan.
- . 1993. "Mental Events as Structuring Causes of Behaviour". En *Mental Causation*, eds. J. Heil y A.R. Mele, 121-36. Oxford: Clarendon Press.
- . 1988. *Explaining Behavior: Reasons in a World of Causes*. Cambridge, Massachusetts: MIT Press.
- Eccles, J.C. 1994. *How the Self Controls Its Brain*. Berlín: Springer.
- . 1986. "Do Mental Events Cause Neural Events Analogously to the Probability Fields of Quantum Mechanics?" *Proceedings of the Royal Society of London B* 227: 411-28.
- Einstein, A. 1905. "Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt". *Annalen der Physik* 322: 132-48.
- Einstein, A., B. Podolsky, y N. Rosen. 1935. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" *Physical Review* 47: 777-80.
- Ellis, G.F.R. 2009. "Top-Down Causation and the Human Brain". En *Downward Causation and the Neurobiology of Free Will*, eds. N. Murphy, G.F.R. Ellis y T. O'Connor, 63-81. Berlín: Springer.
- Escher, M.C. 1992. *The Graphic Work*. Colonia: Taschen.
- Everett, H., III. 1957. "'Relative State' Formulation of Quantum Mechanics". *Reviews of Modern Physics* 29: 454-62.
- Fellous, J-M. 1999. "Neuromodulatory Basis of Emotion". *Neuroscientist* 5: 283-94.

- Feynman, R.P. 1948. "Space-Time Approach to Non-Relativistic Quantum Mechanics". *Reviews of Modern Physics* 20: 367-87.
- Finch, A., y T.A. Warfield. 1998. "The *Mind* Argument and Libertarianism". *Mind* 107: 515-28.
- Fischer, J.M. 2007. "Compatibilism". En *Four Views on Free Will*, por J.M. Fischer, R. Kane, D. Pereboom y M. Vargas, 44-84. Malden, Massachusetts: Blackwell.
- Fischer, J.M., R. Kane, D. Pereboom, y M. Vargas. 2007. *Four Views on Free Will*. Malden, Massachusetts: Blackwell.
- Flegr, J. 2013. "Influence of Latent *Toxoplasma* Infection on Human Personality, Physiology and Morphology: Pros and Cons of the *Toxoplasma*–Human Model in Studying the Manipulation Hypothesis". *The Journal of Experimental Biology* 216: 127-33.
- . 2007. "Effects of *Toxoplasma* on Human Behavior". *Schizophrenia Bulletin* 33: 757-60.
- Flegr, J., Š. Zitková, P. Kodym, y D. Frynta. 1996. "Induction of Changes in Human Behaviour by the Parasitic Protozoan *Toxoplasma gondii*". *Parasitology* 113: 49-54.
- Frankfurt, H.G. 1978. "The Problem of Action". *American Philosophical Quarterly* 15: 157-62.
- . 1971. "Freedom of the Will and the Concept of a Person". *The Journal of Philosophy* 68: 5-20.
- . 1969. "Alternate Possibilities and Moral Responsibility". *The Journal of Philosophy* 66: 829-39.
- Frith, C.D. 2009. "Free Will and Top-Down Control in the Brain". En *Downward Causation and the Neurobiology of Free Will*, eds. N. Murphy, G.F.R. Ellis y T. O'Connor, 199-209. Berlín: Springer.
- Frith, C.D., K. Friston, P.F. Liddle, y R.S.J. Frackowiak. 1991. "Willed Action and the Prefrontal Cortex in Man: A Study with PET". *Proceedings of the Royal Society B: Biological Sciences* 244: 241-6.

- Fuster, J.M. 1995. *Memory in the Cerebral Cortex: An Empirical Approach to Neural Networks in the Human and Nonhuman Primate*. Cambridge, Massachusetts: MIT Press.
- Gasquoine, P.G. 1993. "Alien Hand Sign". *Journal of Clinical and Experimental Neuropsychology* 15: 653-67.
- Ginet, C. 1990. *On Action*. Cambridge: Cambridge University Press.
- Gisin, N. 1989. "Stochastic Quantum Dynamics and Relativity". *Helvetica Physica Acta* 62: 363-71.
- . 1984. "Quantum Measurements and Stochastic Processes". *Physical Review Letters* 52: 1657-60.
- Goldberg, G., y K.K. Bloom. 1990. "The Alien Hand Sign: Localization, Lateralization and Recovery". *American Journal of Physical Medicine & Rehabilitation* 69: 228-38.
- Goldman-Rakic, P.S. 1987. "Motor Control Function of the Prefrontal Cortex". En *Motor Areas of the Cerebral Cortex: Ciba Foundation Symposium 132*, ed. Ciba Foundation Symposium 132, 187-200. Chichester, Reino Unido: Wiley.
- Goldstein, S., D.V. Tausk, R. Tumulka, y N. Zanghì. 2010. "What Does the Free Will Theorem Actually Prove?" *Notices of the AMS* 57: 1451-3.
- Gomes, G. 1999. "Volition and the Readiness Potential". En *The Volitional Brain: Towards a Neuroscience of Free Will*, eds. B. Libet, A. Freeman y K. Sutherland, 59-76. Exeter, Reino Unido: Imprint Academic.
- Grassberger, P. 1989. "Problems in Quantifying Self-Generated Complexity". *Helvetica Physica Acta* 62: 489-511.
- Greene, B. 1999. *The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory*. Nueva York: Norton.
- Greene, J.D., R.B. Sommerville, L.E. Nystrom, J.M. Darley, y J.D. Cohen. 2001. "An fMRI Investigation of Emotional Engagement in Moral Judgment". *Science* 293: 2105-8.

- Gribbin, J. 1985. *In Search of Schrodinger's Cat: Quantum Physics and Reality*. Nueva York: Bantam.
- Haggard, P., y E. Magno. 1999. "Localising Awareness of Action with Transcranial Magnetic Stimulation". *Experimental Brain Research* 127: 102-7.
- Hameroff, S.R. 2012. "How Quantum Brain Biology Can Rescue Conscious Free Will". *Frontiers in Integrative Neuroscience* 6: 93.
- Hameroff, S.R., y R. Penrose. 1996. "Conscious Events as Orchestrated Space-Time Selections". *Journal of Consciousness Studies* 3: 36-53.
- Harbecke, J., y H. Atmanspacher. 2012. "Horizontal and Vertical Determination of Mental and Neural States". *Journal of Theoretical and Philosophical Psychology* 32: 161-79.
- Hawking, S.W. 1988. *A Brief History of Time: From the Big Bang to Black Holes*. Nueva York: Bantam.
- Heisenberg, W. 1958. *Physics and Philosophy: The Revolution in Modern Science*. Londres: George Allen & Unwin.
- . 1927. "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik". *Zeitschrift für Physik A: Hadrons and Nuclei* 43: 172-98.
- Hobbs, J. 1991. "Chaos and Indeterminism". *Canadian Journal of Philosophy* 21: 141-64.
- Hodgson, D. 2002. "Quantum Physics, Consciousness, and Free Will". En *The Oxford Handbook of Free Will*, ed. R. Kane, 85-110. Nueva York: Oxford University Press.
- Hofmann, H.F. 2012. "Complex Joint Probabilities as Expressions of Reversible Transformations in Quantum Mechanics". *New Journal of Physics* 14: 043031.
- Hofstadter, D.R. 2007. *I Am a Strange Loop*. Nueva York: Basic Books.
- . 1979. *Gödel, Escher, Bach: An Eternal Golden Braid*. Nueva York: Basic Books.

- Honderich, T. 2002. "Determinism as True, Both Compatibilism and Incompatibilism as False, and the Real Problem". En *The Oxford Handbook of Free Will*, ed. R. Kane, 461-76. Nueva York: Oxford University Press.
- . 1993. *How Free Are You?: The Determinism Problem*. Oxford: Oxford University Press.
- Hookway, C. 2001. "Epistemic *Akrasia* and Epistemic Virtue". En *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility*, eds. A. Fairweather y L. Zagzebski, 178-99. Nueva York: Oxford University Press.
- Horgan, T. 1991. "Actions, Reasons, and the Explanatory Role of Content". En *Dretske and His Critics*, ed. B. McLaughlin, 73-101. Oxford: Basil Blackwell.
- Huemer, M., y B. Kovitz. 2003. "Causation as Simultaneous and Continuous". *The Philosophical Quarterly* 53: 556-65.
- Huttenlocher, P.R. 1979. "Synaptic Density in Human Frontal Cortex — Developmental Changes and Effects of Aging". *Brain Research* 163: 195-205.
- Huttenlocher, P.R., y A.S. Dabholkar. 1997. "Regional Differences in Synaptogenesis in Human Cerebral Cortex". *The Journal of Comparative Neurology* 387: 167-78.
- Hyder, F., E.A. Phelps, C.J. Wiggins, K.S. Labar, A.M. Blamire, y R.G. Shulman. 1997. "'Willed Action": A Functional MRI Study of the Human Prefrontal Cortex During a Sensorimotor Task". *Proceedings of the National Academy of Sciences of the USA* 94: 6989-94.
- Jahanshahi, M., I.H. Jenkins, R.G. Brown, C.D. Marsden, R.E. Passingham, y D.J. Brooks. 1995. "Self-Initiated versus Externally Triggered Movements: I. An Investigation Using Measurement of Regional Cerebral Blood Flow with PET and Movement-Related Potentials in Normal and Parkinson's Disease Subjects". *Brain* 118: 913-33.

- Johansson, S., y P. Arhem. 1994. "Single-Channel Currents Trigger Action Potentials in Small Cultured Hippocampal Neurons". *Proceedings of the National Academy of Sciences of the USA* 91: 1761-5.
- Jordan, P. 1938. "Die Verstärkertheorie der Organismen in ihrem gegenwärtigen Stand". *Naturwissenschaften* 26: 537-45.
- . 1934. "Quantenphysikalische Bemerkungen zur Biologie und Psychologie". *Erkenntnis* 4: 215-52.
- . 1932. "Die Quantenmechanik und die Grundprobleme der Biologie und Psychologie". *Naturwissenschaften* 20: 815-21.
- Juarrero, A. 2009. "Top-Down Causation and Autonomy in Complex Systems". En *Downward Causation and the Neurobiology of Free Will*, eds. N. Murphy, G.F.R. Ellis y T. O'Connor, 83-102. Berlín: Springer.
- . 1999. *Dynamics in Action: Intentional Behavior as a Complex System*. Cambridge, Massachusetts: MIT Press.
- Kane, R. 2011. "Rethinking Free Will: New Perspectives on an Ancient Problem". En *The Oxford Handbook of Free Will (2nd ed.)*, ed. R. Kane, 381-404. Nueva York: Oxford University Press.
- . 2009. "Free Will and the Dialectic of Selfhood: Can One Make Sense of a Traditional Free Will Requiring Ultimate Responsibility? *Ideas y Valores* 58: 25-43.
- . 2008. "Three Freedoms, Free Will and Self-Formation: A Reply to Levy and Other Critics". En *Essays on Free Will and Moral Responsibility*, eds. N. Trakakis y D. Cohen, 142-62. Newcastle upon Tyne, Reino Unido: Cambridge Scholars.
- . 2007a. "Libertarianism". En *Four Views on Free Will*, por J.M. Fischer, R. Kane, D. Pereboom y M. Vargas, 5-43. Malden, Massachusetts: Blackwell.
- . 2007b. "Response to Fischer, Pereboom, and Vargas". En *Four Views on Free Will*, por J.M. Fischer, R. Kane, D. Pereboom y M. Vargas, 166-83. Malden, Massachusetts: Blackwell.

- . 2006. “Review: Libertarian Accounts of Free Will”. *Mind* 115: 136-42.
- . 2005. *A Contemporary Introduction to Free Will*. Nueva York: Oxford University Press.
- . 2002a. “Introduction: The Contours of Contemporary Free Will Debates”. En *The Oxford Handbook of Free Will*, ed. R. Kane, 3-41. Nueva York: Oxford University Press.
- . 2002b. “Some Neglected Pathways in the Free Will Labyrinth”. En *The Oxford Handbook of Free Will*, ed. R. Kane, 406-37. Nueva York: Oxford University Press.
- . 2000. “The Dual Regress of Free Will and the Role of Alternative Possibilities”. *Philosophical Perspectives* 14: 57-79.
- . 1999. “Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism”. *The Journal of Philosophy* 96: 217-40.
- . 1996. *The Significance of Free Will*. Nueva York: Oxford University Press.
- . 1989. “Two Kinds of Incompatibilism”. *Philosophy and Phenomenological Research* 50: 219-54.
- . 1985. *Free Will and Values*. Albany, Nueva York: State University of New York Press.
- Kelso, J.A.S., y E. Tognoli. 2009. “Toward a Complementary Neuroscience: Metastable Coordination Dynamics of the Brain”. En *Downward Causation and the Neurobiology of Free Will*, eds. N. Murphy, G.F.R. Ellis y T. O'Connor, 103-24. Berlín: Springer.
- Kennard, E.H. 1927. “Zur Quantenmechanik einfacher Bewegungstypen”. *Zeitschrift für Physik* 44: 326-52.
- Kim, J. 2006. “Emergence: Core Ideas and Issues”. *Synthese* 151: 547-59.
- . 1999. “Making Sense of Emergence”. *Philosophical Studies* 95: 3-36.
- . 1993. “The Non-Reductivist's Troubles with Mental Causation”. En *Mental Causation*, eds. J. Heil y A.R. Mele, 189-210. Oxford: Clarendon Press.

- . 1992. "Downward Causation" in Emergentism and Nonreductive Physicalism". En *Emergence or Reduction?: Essays on the Prospects of Nonreductive Physicalism*, eds. A. Beckermann, H. Flohr y J. Kim, 119-38. Berlín: de Gruyter.
- Knobe, J., y S. Nichols. 2011. "Free Will and the Bounds of the Self". En *The Oxford Handbook of Free Will (2nd ed.)*, ed. R. Kane, 530-54. Nueva York: Oxford University Press.
- Kochen, S., y E.P. Specker. 1967. "The Problem of Hidden Variables in Quantum Mechanics". *Journal of Mathematics and Mechanics* 17: 59-87.
- Koenigs, M., L. Young, R. Adolphs, D. Tranel, F. Cushman, M. Hauser, y A. Damasio. 2007. "Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgements". *Nature* 446: 908-11.
- Koons, R.C., y T.H. Pickavance. 2015. *Metaphysics: The Fundamentals*. Chichester, Reino Unido: Wiley Blackwell.
- Kornhuber, H.H., y L. Deecke. 1965. "Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale". *Pflügers Archiv: European Journal of Physiology* 284: 1-17.
- Kurzban, R., P. DeScioli, y E. O'Brien. 2007. "Audience Effects on Moralistic Punishment". *Evolution & Human Behavior* 28: 75-84.
- Lamb, J.W. 1993. "Evaluative Compatibilism and the Principle of Alternate Possibilities". *The Journal of Philosophy* 90: 517-27.
- . 1977. "On a Proof of Incompatibilism". *The Philosophical Review* 86: 20-35.
- Lau, H.C., R.D. Rogers, y R.E. Passingham. 2007. "Manipulating the Experienced Onset of Intention after Action Execution". *Journal of Cognitive Neuroscience* 19: 81-90.
- Lhermitte, F. 1983. "Utilization Behaviour' and Its Relation to Lesions of the Frontal Lobes". *Brain* 106: 237-55.

- Libet, B. 2004. *Mind Time: The Temporal Factor in Consciousness*. Cambridge, Massachusetts: Harvard University Press.
- . 1999. “Do We Have Free Will?”. En *The Volitional Brain: Towards a Neuroscience of Free Will*, eds. B. Libet, A. Freeman y K. Sutherland, 47-57. Exeter, Reino Unido: Imprint Academic.
- . 1996. “Neural Time Factors in Conscious and Unconscious Mental Functions”. En *Toward a Science of Consciousness: The First Tucson Discussions and Debates*, eds. S.R. Hameroff, A.W. Kaszniak y A.C. Scott, 337-48. Cambridge, Massachusetts: MIT Press.
- . 1994. “A Testable Field Theory of Mind-Brain Interaction”. *Journal of Consciousness Studies* 1: 119-26.
- . 1993. “The Neural Time Factor in Conscious and Unconscious Events”. En *Experimental and Theoretical Studies of Consciousness: Ciba Foundation Symposium 174*, ed. Ciba Foundation Symposium 174, 123-46. Chichester, Reino Unido: Wiley.
- . 1989. “Conscious Subjective Experience vs. Unconscious Mental Functions: A Theory of the Cerebral Processes Involved”. En *Models of Brain Function*, ed. R.M.J. Cotterill, 68-79. Nueva York: Cambridge University Press.
- . 1985. “Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action”. *Behavioral and Brain Sciences* 8: 529-66.
- Libet, B., C.A. Gleason, E.W. Wright, y D.K. Pearl. 1983. Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential): The Unconscious Initiation of a Freely Voluntary Act”. *Brain* 106: 623-42.
- Libet, B., D.K. Pearl, D.E. Morledge, C.A. Gleason, Y. Hosobuchi, y N.M. Barbaro. 1991. “Control of the Transition from Sensory Detection to Sensory Awareness in Man by the Duration of a Thalamic Stimulus: The Cerebral ‘Time-on’ Factor. *Brain* 114: 1731-57.

- Libet, B., E.W. Wright, B. Feinstein Jr., y D.K. Pearl. 1979. "Subjective Referral of the Timing for a Conscious Sensory Experience: A Functional Role for the Somatosensory Specific Projection System in Man". *Brain* 102: 193-224.
- Libet, B., E.W. Wright, y C.A. Gleason. 1983. "Preparation- or Intention-to-Act, in Relation to Pre-Event Potentials Recorded at the Vertex". *Electroencephalography and Clinical Neurophysiology*, 56: 367-72.
- . 1982. "Readiness-Potentials Preceding Unrestricted 'Spontaneous' vs. Pre-Planned Voluntary Acts". *Electroencephalography and Clinical Neurophysiology* 54: 322-35.
- Liebovitch, L.S., y T.I. Toth. 1991. "A Model of Ion Channel Kinetics Using Deterministic Chaotic Rather than Stochastic Processes". *Journal of Theoretical Biology* 148: 243-67.
- Lockwood, M. 1989. *Mind, Brain and the Quantum: The Compound 'I'*. Oxford: Blackwell.
- Lombardi, O., M. Castagnino, y J.S. Ardenghi. 2009. "Mecánica cuántica: Interpretación e invariancia". *Theoria* 64: 5-28.
- Lucas, J.R. 1961. "Minds, Machines and Gödel". *Philosophy* 36: 112-27.
- Mele, A.R. 2011. "Free Will and Science". En *The Oxford Handbook of Free Will (2nd ed.)*, ed. R. Kane, 499-514. Nueva York: Oxford University Press.
- . 2009. *Effective Intentions: The Power of Conscious Will*. Nueva York: Oxford University Press.
- . 2002. "Autonomy, Self-Control, and Weakness of Will". En *The Oxford Handbook of Free Will*, ed. R. Kane, 529-48. Nueva York: Oxford University Press.
- . 1999. "Kane, Luck, and the Significance of Free Will". *Philosophical Explorations* 2: 96-104.
- . 1998. "Review: The Significance of Free Will". *The Journal of Philosophy*, 95: 581-4.

- Mele, A.R., y D. Robb. 1998. "Rescuing Frankfurt-Style Cases". *The Philosophical Review* 107: 97-112.
- Miller, E.K., y J.D. Cohen. 2001. "An Integrative Theory of Prefrontal Cortex Function". *Annual Review of Neuroscience* 24: 167-202.
- Miller, E.K., y J. Wallis. 2008. "The Prefrontal Cortex and Executive Brain Functions". En *Fundamental Neuroscience (3rd ed.)*, eds. L.R. Squire, D. Berg, F.E. Bloom, S. du Lac, A. Ghosh y N.C. Spitzer, 1199-222. Burlington, Massachusetts: Academic Press.
- Morden, M. 1990. "Free Will, Self-Causation, and Strange Loops". *Australasian Journal of Philosophy* 68: 59-73.
- Moya, C.J. 2012. "Libertad, responsabilidad moral y neurociencia". *Pasajes: Revista de Pensamiento Contemporáneo* 37: 109-18.
- . 2011. "Mind, Brain, and Downward Causation". En *Moral Behavior and Free Will: A Neurobiological and Philosophical Approach*, eds. J.J. Sanguinetti, A. Acerbi y J.A. Lombo, 185-200. Morolo, Italia: IF Press.
- . 2006a. *Filosofía de la mente (2ª ed.)*. Valencia: PUV.
- . 2006b. *Moral Responsibility: The Ways of Scepticism*. Nueva York: Routledge.
- . 2004. "Donald Davidson y las ciencias sociales". *Pasajes: Revista de Pensamiento Contemporáneo* 14: 59-67.
- . 1998a. "Justificación, causalidad y acción intencional". *Theoria* 13: 349-65.
- . 1998b. "Reason and Causation in Davidson's Theory of Action Explanation". *Crítica* 30: 29-43.
- . 1994. "Las emociones y la naturalización de la intencionalidad". *Anales del Seminario de Metafísica* 28: 227-55.
- Murphy, N. 2011. "Avoiding Neurobiological Reductionism: The Role of Downward Causation in Complex Systems". En *Moral Behavior and Free Will: A Neurobiological and Philosophical Approach*, eds. J.J. Sanguinetti, A. Acerbi y J.A. Lombo, 201-22. Morolo, Italia: IF Press.

- Murphy, N., y W.S. Brown. 2007. *Did My Neurons Make Me Do It?: Philosophical and Neurobiological Perspectives on Moral Responsibility and Free Will*. Oxford: Oxford University Press.
- Nahmias, E. 2011. "Intuitions about Free Will, Determinism, and Bypassing". En *The Oxford Handbook of Free Will (2nd ed.)*, ed. R. Kane, 555-76. Nueva York: Oxford University Press.
- . 2002. "When Consciousness Matters: A Critical Review of Daniel Wegner's *The Illusion of Conscious Will*". *Philosophical Psychology* 15: 527-41.
- Newsome, W.T. 2009. "Human Freedom and "Emergence"". En *Downward Causation and the Neurobiology of Free Will*, eds. N. Murphy, G.F.R. Ellis y T. O'Connor, 53-62. Berlín: Springer.
- Pacherie, E. 2006. "Toward a Dynamic Theory of Intentions". En *Does Consciousness Cause Behavior?*, eds. S. Pocket, W.P. Banks y S. Gallagher, 145-67. Cambridge, Massachusetts: MIT Press.
- Paolicelli, R.C., G. Bolasco, F. Pagani, L. Maggi, M. Scianni, P. Panzanelli, M. Giustetto, T.A. Ferreira, E. Guiducci, L. Dumas, D. Ragozzino, y C.T. Gross. 2011. "Synaptic Pruning by Microglia Is Necessary for Normal Brain Development". *Science* 333: 1456-8.
- Papineau, D. 2002. *Thinking about Consciousness*. Nueva York: Oxford University Press.
- Penrose, R. 1994. *Shadows of the Mind: A Search for the Missing Science of Consciousness*. Oxford: Oxford University Press.
- . 1989. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. Oxford: Oxford University Press.
- Pereboom, D. 2007. "Hard Incompatibilism". En *Four Views on Free Will*, por J.M. Fischer, R. Kane, D. Pereboom y M. Vargas, 85-125. Malden, Massachusetts: Blackwell.
- . 2001. *Living Without Free Will*. Cambridge: Cambridge University Press.

- Pert, C.B., M.R. Ruff, R.J. Weber, y M. Herkenham. 1985. "Neuropeptides and Their Receptors: A Psychosomatic Network". *The Journal of Immunology* 135: 820-26.
- Petanjek, Z., M. Judaš, G. Šimic, M.R. Rašin, H.B.M. Uylings, P. Rakic, y I. Kostovic. 2011. "Extraordinary Neoteny of Synaptic Spines in the Human Prefrontal Cortex". *Proceedings of the National Academy of Sciences of the USA* 108: 13281-6.
- Petrosky, T., y I. Prigogine. 1997. "The Extension of Classical Dynamics for Unstable Hamiltonian Systems". *Computers & Mathematics with Applications* 34: 1-44.
- . 1996. "Poincaré Resonances and the Extension of Classical Dynamics". *Chaos, Solitons & Fractals* 7: 441-97.
- Planck, M. 1900. "Zur Theorie des Gesetzes der Energieverteilung im Normalspectrum". *Verhandlungen Der Deutschen Physikalischen Gesellschaft* 2: 237-45.
- Prigogine, I. 2003. "Is Future Given? Changes in Our Description of Nature". En *Is Future Given?*, por I. Prigogine, I.E. Antoniou, T. Christidis, I. Zisis y M. Adamidou, 7-20. Singapur: World Scientific.
- . 1997. *The End of Certainty: Time, Chaos, and the New Laws of Nature*. Nueva York: Free Press.
- Quiroga, R.Q., L. Reddy, G. Kreiman, C. Koch, y I. Fried. 2005. "Invariant Visual Representation by Single Neurons in the Human Brain". *Nature* 435: 1102-7.
- Sadiq, M., y F. Ghafoor. 2010. "An Alternative Realization of 't Hooft's Quantum Determinism". *Physics Letters A* 375: 11-13.
- Schardt, D.M., S. Erk, C. Nüsser, M.M. Nöthen, S. Cichon, M. Rietschel, J. Treutlein, T. Goschke, y H. Walter. 2010. "Volition Diminishes Genetically Mediated Amygdala Hyperreactivity". *NeuroImage* 53: 943-51.
- Schrödinger, E. 1936. "Die gegenwärtige Situation in der Quantenmechanik". *Naturwissenschaften* 23: 807-12.

- . 1926a. “Quantisierung als Eigenwertproblem (Erste Mitteilung)”. *Annalen der Physik* 384: 361-76.
- . 1926b. “Quantisierung als Eigenwertproblem (Zweite Mitteilung)”. *Annalen der Physik* 384: 489-527.
- . 1926c. “Quantisierung als Eigenwertproblem (Dritte Mitteilung)”. *Annalen der Physik* 385: 437-90.
- . 1926d. “Quantisierung als Eigenwertproblem (Vierte Mitteilung)”. *Annalen der Physik* 386: 109-39.
- Schultz, W. 1999. “The Primate Basal Ganglia and the Voluntary Control of Behaviour”. En *The Volitional Brain: Towards a Neuroscience of Free Will*, eds. B. Libet, A. Freeman y K. Sutherland, 31-45. Exeter, Reino Unido: Imprint Academic.
- Schwartz, J.M. 1999. “A Role for Volition and Attention in the Generation of New Brain Circuitry: Toward a Neurobiology of Mental Force”. En *The Volitional Brain: Towards a Neuroscience of Free Will*, eds. B. Libet, A. Freeman y K. Sutherland, 115-42. Exeter, Reino Unido: Imprint Academic.
- Smilansky, S. 2000. *Free Will and Illusion*. Oxford: Clarendon Press.
- Sober, E. 1984. “Two Concepts of Cause”. *PSA* 1984: 405-24.
- Soon, C.S., M. Brass, H-J Heinze, y J-D Haynes. 2008. “Unconscious Determinants of Free Decisions in the Human Brain”. *Nature Neuroscience* 11: 543-5.
- Spalding, K.L., O. Bergmann, K. Alkass, S. Bernard, M. Salehpour, H.B. Huttner, E. Boström, I. Westerlund, C. Vial, B.A. Buchholz, G. Possnert, D.C. Mash, H. Druid, y J. Frisé. 2013. “Dynamics of Hippocampal Neurogenesis in Adult Humans”. *Cell* 153: 1219-27.
- Spence, S.A., D.J. Brooks, S.R. Hirsch, P.F. Liddle, J. Meehan, y P.M. Grasby. 1997. “A PET Study of Voluntary Movement in Schizophrenic Patients Experiencing Passivity Phenomena (Delusions of Alien Control)”. *Brain* 120: 1997-2011.

- Spence, S.A., y C.D. Frith. 1999. "Towards a Functional Anatomy of Volition". En *The Volitional Brain: Towards a Neuroscience of Free Will*, eds. B. Libet, A. Freeman y K. Sutherland, 11-29. Exeter, Reino Unido: Imprint Academic.
- Squire, L.R., D. Berg, F.E. Bloom, S. du Lac, A. Ghosh, y N.C. Spitzer, eds. 2008. *Fundamental Neuroscience (3<sup>rd</sup> ed.)*. Burlington, Massachusetts: Academic Press.
- Stapp, H.P. 2007. *Mindful Universe: Quantum Mechanics and the Participating Observer*. Berlín: Springer.
- . 1999. "Attention, Intention, and Will in Quantum Mechanics". En *The Volitional Brain: Towards a Neuroscience of Free Will*, eds. B. Libet, A. Freeman y K. Sutherland, 143-64. Exeter, Reino Unido: Imprint Academic.
- . 1998. "Pragmatic Approach to Consciousness". En *Brain and Values: Is a Biological Science of Values Possible*, ed. K.H. Pribram, 237-48. Mahwah, Nueva Jersey: Lawrence Erlbaum.
- . 1993. *Mind, Matter, and Quantum Mechanics*. Berlín: Springer.
- Strawson, G. 1994. "The Impossibility of Moral Responsibility". *Philosophical Studies* 75: 5-24.
- . 1986. *Freedom and Belief*. Oxford: Oxford University Press.
- Strawson, P.F. 1962. "Freedom and Resentment". *Proceedings of the British Academy* 48: 1-25.
- 't Hooft, G. 2007. "A Mathematical Theory for Deterministic Quantum Mechanics". *Journal of Physics: Conference Series* 67: 012015.
- Tapia, J.C., y J.W. Lichtman. 2008. "Synapse Elimination". En *Fundamental Neuroscience (3<sup>rd</sup> ed.)*, eds. L.R. Squire, D. Berg, F.E. Bloom, S. du Lac, A. Ghosh y N.C. Spitzer, 469-90. Burlington, Massachusetts: Academic Press.
- Taylor, C., y D.C. Dennett. 2002. "Who's Afraid of Determinism? Rethinking Causes and Possibilities". En *The Oxford Handbook of Free Will*, ed. R. Kane, 257-77. Nueva York: Oxford University Press.

- Tegmark, M. 2000. "Importance of Quantum Decoherence in Brain Processes". *Physical Review E* 61: 4194-206.
- Trevena, J., y J. Miller. 2010. "Brain Preparation before a Voluntary Action: Evidence Against Unconscious Movement Initiation". *Consciousness and Cognition* 19: 447-56.
- Tumulka, R. 2007. "Comment on "The Free Will Theorem"". *Foundations of Physics* 37: 186-97.
- Van Gulick, R. 1993. "Who's in Charge Here? And Who's Doing All the Work?". En *Mental Causation*, eds. J. Heil y A.R. Mele, 233-56. Oxford: Clarendon Press.
- van Inwagen, P. 2000. "Free Will Remains a Mystery". *Philosophical Perspectives* 14: 1-19.
- . 1983. *An Essay on Free Will*. Oxford: Clarendon Press.
- . 1975. "The Incompatibility of Free Will and Determinism". *Philosophical Studies* 27: 185-99.
- Vargas, M. 2009. "Revisionism about Free Will: A Statement & Defense". *Philosophical Studies* 144: 45-62.
- . 2007. "Revisionism". En *Four Views on Free Will*, por J.M. Fischer, R. Kane, D. Pereboom y M. Vargas, 126-65. Malden, Massachusetts: Blackwell.
- Vicente, A. 2000. "La teoría dretskeana de la causación mental ante la explicación psicológica". *Éndoxa* 13: 9-14.
- Vohs, K.D., y J.W. Schooler. 2008. "The Value of Believing in Free Will: Encouraging a Belief in Determinism Increases Cheating". *Psychological Science* 19: 49-54.
- von Neumann, J. 1955. *Mathematical Foundations of Quantum Mechanics*. Princeton: Princeton University Press.
- Walter, H. 2011. "Contributions of Neuroscience to the Free Will Debate: From Random Movement to Intelligible Action". En *The Oxford Handbook*

- of *Free Will* (2nd ed.), ed. R. Kane, 515-29. Nueva York: Oxford University Press.
- . 2002. “Neurophilosophy of Free Will”. En *The Oxford Handbook of Free Will*, ed. R. Kane, 565-76. Nueva York: Oxford University Press.
- . 2001. *Neurophilosophy of Free Will: From Libertarian Illusions to a Concept of Natural Autonomy*. Cambridge, Massachusetts: MIT Press.
- Wallace, R.J. 1994. *Responsibility and the Moral Sentiments*. Cambridge, Massachusetts: Harvard University Press.
- Waller, B.N. 1990. *Freedom Without Responsibility*. Filadelfia: Temple University Press.
- Warfield, T.A. 1996. “Determinism and Moral Responsibility Are Incompatible”. *Philosophical Topics* 24: 215-26.
- Watson, G. 1975. “Free Agency”. *The Journal of Philosophy* 72: 205-20.
- Weber, M. 2005. “Indeterminism in Neurobiology”. *Philosophy of Science* 72: 663-74.
- Wegner, D.M. 2004. “Précis of *The Illusion of Conscious Will*”. *Behavioral and Brain Sciences* 27: 649-59.
- . 2002. *The Illusion of Conscious Will*. Cambridge, Massachusetts: MIT Press.
- Widerker, D. 1987. “On an Argument for Incompatibilism”. *Analysis* 47: 37-41.
- Winch, P. 1963. *The Idea of a Social Science and Its Relation to Philosophy*. 3ª impresión (con algunas correcciones). Londres: Routledge & Kegan Paul.
- Wolf, S. 1990. *Freedom Within Reason*. Nueva York: Oxford University Press.
- Zurek, W.H. 1991. “Decoherence and the Transition from Quantum to Classical”. *Physics Today* 44: 36-44.
- . 1986. “Reduction of the Wave Packet and Environment-Induced Superselection”. *Annals of the New York Academy of Sciences* 480: 89-97.
- . 1982. “Environment-Induced Superselection Rules”. *Physical Review D* 26: 1862-80.

———. 1981. “Pointer Basis of Quantum Apparatus: Into What Mixture Does the Wave Packet Collapse?” *Physical Review D* 24: 1516-25.

