

La neurofilosofía como punto de encuentro entre filosofía y neurociencias

ALICIA RODRÍGUEZ SERÓN
Universidad de Málaga

RESUMEN

Los progresos tecnológicos y teóricos alcanzados por las neurociencias en estos últimos veinticinco años, y sus implicaciones para la filosofía, han renovado las posibilidades de aproximación entre ambas, propiciando el desarrollo de un nuevo ámbito de estudio dentro de la filosofía: la neurofilosofía. El presente trabajo se propone sugerir cómo este campo emergente podría ser concebido como un punto de encuentro para la reflexión conjunta y la interacción entre filosofía y neurociencias, no marcado necesariamente por la absorción de la primera por la segunda.

PALABRAS CLAVE

NEUROFILOSOFÍA, FILOSOFÍA, NEUROCIENCIAS, CIENCIAS COGNITIVAS

ABSTRACT

The technological and theoretical advances in neurosciences in the past twenty-five years, as well as their implications for philosophy have opened promising perspectives for the renewal of approach between both, thus encouraging the development of a arising field within philosophy: neurophilosophy. The purpose of this paper is to suggest how this new field study could be considered to be a point of union for the joint reflection and interaction between philosophy and neurosciences, without the first being absorbed by the second.

KEY WORDS

NEUROPHILOSOPHY, PHILOSOPHY, NEUROSCIENCES, COGNITIVE SCIENCES

EL ENORME PROGRESO ALCANZADO POR LAS CIENCIAS COGNITIVAS en estos últimos años ha renovado las posibilidades de aproximación entre neurociencias y filosofía de una manera inusitada, favoreciendo el desarrollo de un nuevo campo dentro de la filosofía denominado neurofilosofía. El término fue populari-

zado por Patricia Smith Churchland en 1986, en su obra *Neurophilosophy: toward a unified science of the mind-brain*, para designar una concepción eliminativa de los procesos mentales que vendría a negar las diferencias entre la filosofía y las neurociencias y en donde la primera acabaría siendo absorbida por las segundas. Sin embargo, el neologismo no tiene por qué remitir sistemáticamente al sentido definido por P. S. Churchland. La neurofilosofía puede ser concebida como un punto de encuentro fecundo entre filosofía y neurociencias donde las relaciones entre ambas no tienen que estar marcadas necesariamente por la eliminación o la absorción.

El encuentro no debe suponer dominio y subordinación de una de las partes, sino colaboración, interacción y reflexión conjunta; no debe hacer de la filosofía una rama de las neurociencias, sino un lugar intermedio en el que puedan plantearse ciertas condiciones de posibilidad de los fenómenos mentales, en un nivel relativamente abstracto; una vez formuladas tales condiciones, éstas deberán poder ser confrontadas con las investigaciones experimentales en neurociencias. Esta manera de entender las relaciones vuelve a replantear la vieja cuestión, de carácter más general, de la conexión y fronteras entre ciencia y filosofía.

I. ¿PESIMISMO U OPTIMISMO?

A juzgar por la proliferación de congresos y publicaciones interdisciplinarias, hay motivos para ser optimistas sobre el futuro de las relaciones entre filosofía y neurociencias. Pero por muy optimistas que seamos, no podemos pasar por alto la persistencia de un cierto pesimismo que sigue latente y que, como es lógico, puede provenir de cada ámbito. Por parte de los neurocientíficos, cabe preguntarse si ciertos análisis filosóficos referidos a los procesos mentales –suponiendo que fuese necesario tenerlos en cuenta– contribuirán realmente a la obtención de verdades *acerca de la cognición humana*. Por parte de los filósofos, cabe preguntarse si las verdades obtenidas acerca del funcionamiento y de la estructura de la cognición –suponiendo que fuese realmente necesario tenerlas en cuenta– serán en realidad verdades *filosóficas*.

Es posible compartir semejante pesimismo y sin embargo no defender la idea de que *todas* las verdades filosóficas son necesariamente verdades conceptuales *a priori*, sustraídas por definición a todo control experimental. Y a la inversa, tampoco es necesario sostener, como hacen numerosos filósofos «naturalistas», que en realidad sólo existe una diferencia de grado entre las teorías científicas y las teorías filosóficas y que, por tanto, no habría en principio distinción entre verdades de tipo empírico y verdades de tipo conceptual o *a priori*. Dicho de otra manera, cabe sostener que, aunque haya en filosofía de la mente un cierto número de tesis y conceptos acerca de la cognición que no pueden ser

sometidos a pruebas empíricas, sí existe lo que podríamos llamar una zona intermedia de conceptos que se sitúan en la interfaz entre filosofía y neurociencias cognitivas y que permiten responder de manera razonablemente optimista a las cuestiones antes planteadas. Esta zona que podría denominarse «neurofilosofía», aparentemente sospechosa, no tendría por qué provocar de entrada el rechazo de neurocientíficos y filósofos. Son numerosos los ámbitos en los que existen puntos de encuentro fecundos entre una argumentación de tipo filosófico y las investigaciones en neurociencias. En la actualidad, los más discutidos son la teoría de la mente, la representación y la planificación de acciones. Me limitaré aquí a exponer algunos ejemplos: el problema de la relación mente-cuerpo, la unidad de la consciencia y la naturaleza de la representación mental.

II. RELACIONES MENTE-CUERPO

Sin duda el problema de la relación entre la mente y el cuerpo es con diferencia *el* problema de la filosofía de la mente. Son múltiples las formulaciones que se han ofrecido de él, pero quizá una buena manera de presentarlo de forma sintética consistiría en mostrar cómo las tres proposiciones siguientes, aunque se contradigan mutuamente, parecen plausibles:

- 1) Los hechos mentales son distintos de los hechos físicos,
- 2) Los hechos mentales tienen efectos sobre los hechos físicos,
- 3) Y, sin embargo, los hechos físicos son por sí solos suficientes para causar efectos físicos.

En efecto, aunque reconozcamos que nuestro cerebro interviene en nuestros procesos mentales y que éstos de alguna manera son debidos a aquél, solemos mostrar claras reticencias a la hora de sostener que las propiedades mentales o los hechos mentales son idénticos a propiedades o hechos físicos. Un fisicalista consecuente –aun estando dispuesto a admitir que los hechos mentales son distintos a los hechos físicos– se resistiría a ir más lejos y afirmar que las propiedades o hechos mentales tienen un poder causal sobre las propiedades o hechos físicos, ya que si lo hiciera y reconociera que aquéllos tienen realmente una eficacia causal, no podría defender el fisicalismo sin caer en un dualismo interaccionista. Podría optar por hacer de las propiedades mentales epifenómenos de las propiedades físicas. Pero entonces ¿de qué le serviría reconocer la existencia de los hechos mentales, si a continuación les niega toda eficacia?

Parece por tanto que la mejor manera de preservar el fisicalismo consistiría en negarse a distinguir los acontecimientos, procesos o estados mentales de

los acontecimientos, procesos o estados físicos. La causalidad seguiría siendo entonces un proceso físico sujeto a acontecimientos físicos y no mentales; de lo contrario los hechos mentales representarían algo atípico dentro del mundo físico. Pero lo que no puede negar el fisicalismo es la existencia de experiencias conscientes; sabemos que nuestros deseos, nuestras creencias causan nuestras acciones y nuestros movimientos corporales. Estas tres tesis resultan difíciles de reconciliar y suponen una interpretación del problema mente-cuerpo en términos de causalidad. En función de la aceptación o rechazo de cada una de ellas, se obtendrán posturas diversas.

Como sugiere Tim van Gelder (1998), un dualista interaccionista puede limitarse a abandonar 3); un paralelista negará 2); un fisicalista fuerte o un teórico de la identidad mente-cerebro negará 1). En filosofía como en la vida no se puede tener todo. Sin embargo, cualquier teoría que fuera capaz de reconciliar 1)-3) recibiría una buena acogida, ya que estas intuiciones están muy enraizadas en nosotros. Cabe así mantener una postura intermedia, como la defendida por las distintas versiones contemporáneas de funcionalismo: se admitiría que las propiedades mentales son propiedades funcionales «realizadas» de maneras distintas en las estructuras físicas. También es ésta la posición conciliadora sostenida por autores como Davidson (1980), cuando afirma que, aunque los hechos mentales sean idénticos a los hechos físicos, no podemos albergar la esperanza de explicar de manera completa los primeros mediante los segundos. Tales autores abrazan formas no reduccionistas de fisicalismo, pero se siguen considerando fisicalistas o materialistas en la medida en que defienden una relación especial de dependencia de los hechos mentales con respecto de los hechos físicos, una dependencia no reductiva. Se trata de la tan discutida relación de sobreveniencia, según la cual los hechos mentales sobrevenen sobre los hechos físicos.

El término «sobreveniencia» fue acuñado por G. E. Moore para describir una relación de dependencia entre propiedades morales y propiedades no morales, según la cual es imposible que dos objetos coincidan en todas sus características naturales (descriptivas) y difieran en sus características no naturales (evaluativas). Posteriormente fue adoptado por Davidson para describir la imposibilidad de que dos objetos físicamente idénticos sean diferentes en su aspecto mental (cf. Davidson, 1980, p. 214). La sobreveniencia pasó así a ser interpretada como una relación de dependencia entre dos conjuntos de propiedades, las propiedades sobrevenientes (mentales) y sus propiedades de «base» o subvenientes (físicas). Su fórmula general podría ser la siguiente: todo cambio mental depende de o covaría con un cambio físico, aunque no sea cierto que las mismas diferencias mentales covaríen con o dependan de las mismas diferencias físicas. La sobreveniencia mente-cuerpo puede ser formulada de modo equivalente: si un organismo instancia una propiedad M (como el do-

lor), en ese momento tiene alguna propiedad física P sobre la cual sobreviene M, en el sentido de que si el organismo tiene M entonces tiene P. Si en algún momento experimentamos un dolor, es que hemos debido instanciar una determinada propiedad física (presumiblemente una propiedad neural) sobre la cual sobreviene en ese momento el dolor (Kim, 1993). Dicho de otro modo, ninguna propiedad mental puede ser instanciada en un organismo a menos que este organismo instancie alguna propiedad física adecuada que le sirva de base.

Podría pensarse que la relación de sobrevenida de lo mental sobre lo físico supone la reducción de lo mental a lo físico, sin embargo no es del todo así. La sobrevenida caracteriza un concepto débil de reducción. El hecho de que afirmemos que los acontecimientos mentales se reducen *ontológicamente* a acontecimientos físicos no implica que las propiedades mentales se tengan que reducir *definicionalmente* a las propiedades físicas¹.

Buena parte de la literatura reciente en filosofía de la mente se ha dedicado a explorar una versión u otra de estas posiciones, en particular las del fisicalismo débil y las distintas variedades de sobrevenida de lo mental sobre lo físico. Todas ellas tienen en común el tratar de conjugar la idea de que los acontecimientos mentales son idénticos a los acontecimientos físicos, con la idea de que las propiedades mentales y físicas de los acontecimientos son irreductiblemente distintas. La sobrevenida de las propiedades mentales sobre las propiedades físicas es, a juicio de muchos, la forma más adecuada de plantear la relación entre la mente y el cuerpo, aunque resulte difícil entender cómo puede ser algo más que una simple relación de covarianza entre propiedades físicas y mentales –en el sentido en que la indiscernibilidad de dos propiedades físicas implica la indiscernibilidad de dos propiedades mentales– y sin embargo no implicar la reducción de las segundas a las primeras. La idea de sobrevenida es, sin ningún género de duda, relevante. Nadie cuestionaría su importancia como herramienta metodológica indispensable para discutir una cuestión tan fundamental en filosofía de las ciencias biológicas como es la

1 Se pueden distinguir así al menos dos sentidos de sobrevenida. Un sentido fuerte, según el cual si una propiedad mental M sobreviene sobre una propiedad física P, entonces necesariamente todo lo que es P es M y un sentido débil: si M sobreviene sobre P, la relación entre M y P es la relación de un condicional material («si P entonces M»).

En este segundo sentido, la relación de sobrevenida no es de estricta necesidad. Existen, por tanto, grados de fuerza en la reducción y en la dependencia entre lo mental y lo físico.

Son varios los autores que han distinguido estas dos versiones de sobrevenida. Brian McLaughlin y J. Kim atribuyen a Davidson la versión débil. Pero éstas no son las únicas, se habla también de una sobrevenida global y de una sobrevenida local. Cf. E. LePore y B. McLaughlin, 1985 y J. Kim, 1993.

naturaleza de la reducción y de las variedades más o menos fuertes de reduccionismo. Mantiene además estrechos lazos con las metafísicas emergentistas de finales del siglo XIX. Pero no parece que nos permita dar respuesta al problema mente-cuerpo.

Es cierto que la noción de sobrevenida pretende hacerse eco de una de nuestras intuiciones más comunes acerca de este problema, a saber, la de una dependencia asimétrica entre lo mental y lo físico, a pesar de la no-reducción². Es, por ello, esencial cuando se trata de evaluar la relación entre filosofía y neurofisiología, por ejemplo. Pero, este problema, cuando es tratado desde la relación de sobrevenida, atiende a la ontología de las propiedades mentales y de las propiedades físicas así como a sus condiciones abstractas de covariación, y por esta razón es un problema de naturaleza exclusivamente lógica. Aquí lo que se discute es si una determinada variedad de sobrevenida más fuerte implica alguna otra, más débil, y tales cuestiones tienen que ver sobre todo con la manera en que se definen nociones como propiedad, acontecimiento, instanciación de una propiedad, etc. Las relaciones que se establecen son aquí puramente formales, prueba de ello es que la relación de sobrevenida puede también referirse a la relación entre propiedades morales y propiedades naturales. En este sentido, la noción de sobrevenida no tiene un estatuto explicativo, sino únicamente un estatuto ontológico³.

La cuestión de la sobrevenida de lo mental sobre lo físico sólo podría perder este carácter formal si, en neuropsicología, consideramos situaciones efectivas de covariación de propiedades mentales con respecto a propiedades neurológicas. El método de «sustracción», utilizado especialmente en

2 Del hecho de que B sobrevenga sobre A no se sigue que A sobrevenga sobre B. Quiere decirse que cuando afirmamos que las descripciones mentales sobrevienen sobre las descripciones físicas, hay al menos un sentido en que el carácter físico de un acontecimiento es un rasgo más básico que su carácter mental. La sobrevenida permite la existencia de un universo, totalmente idéntico al nuestro desde el punto de vista físico, en el que no hubiera ningún acontecimiento mental. Pero no supone que pueda existir un universo en el cual hubiera características mentales sin base física, por ejemplo espíritus desencarnados.

3 La tesis de la sobrevenida con respecto a lo físico es una tesis ontológica y no una tesis epistemológica. No implica la reductibilidad de los hechos sobrevenientes sobre el nivel físico a los subvenientes bajo ese nivel. En otras palabras, no implica la identidad de las propiedades no físicas y las propiedades físicas, en el sentido del reduccionismo. Es una opción que podemos tomar, pero que no es obligatoria. La sobrevenida con respecto a lo físico tampoco implica la explicabilidad física. La mejor explicación de un hecho físico no es necesariamente una explicación que pertenezca a la física. Por ejemplo, el hecho de que el coche de mi vecino esté en la puerta de mi casa no se suele explicar en términos físicos, sino por el hecho de que él decidió aparcarlo allí. La teoría física logra explicar la posición del coche físicamente, pero esto no implica que esta explicación sea la explicación más natural que podamos dar.

imágenes cerebrales funcionales, podría servirnos para ilustrar tales situaciones. Se trata de un método que, en general, permite poner en evidencia las variaciones de flujo sanguíneo producidas durante determinadas estimulaciones sensoriales. Consiste en establecer dos condiciones experimentales que, idealmente, son en todo idénticas salvo en la etapa de tratamiento que se desea estudiar. A continuación se sustraen, punto por punto, las medidas fisiológicas efectuadas durante estas dos tareas. La sustracción permite aislar la contribución de las regiones cerebrales implicadas en la función cognitiva estudiada. Este método es muy general y puede ser utilizado tanto para medir el tiempo de reacción como para obtener imágenes tomográficas del flujo sanguíneo cerebral o para estudiar la actividad electromagnética del cerebro (Stanislas Dehaene, 1997). Por ejemplo, si suponemos que una determinada función cognitiva *F* «realiza» la misma propiedad mental (por ejemplo, una tarea de memoria) y deseamos saber cómo esta propiedad covaría con (sobreviene sobre) el flujo sanguíneo o la actividad electromagnética del cerebro, establecemos unas condiciones experimentales en todo idénticas –salvo en lo que se refiere a la etapa de tratamiento memorial que se quiere estudiar– y sustraemos luego las medidas fisiológicas efectuadas en ambas tareas. Podemos decir, entonces, que la propiedad mental sobreviene sobre estas propiedades físicas. Nos encontraríamos ahora, a diferencia del caso precedente, ante una aplicación más local del concepto de sobrevenida. Se ha podido establecer la relación de sobrevenida porque se la ha circunscrito a un entorno menos general.

Pues bien, de la misma manera, cabría suponer que lo que en el ámbito de la filosofía denominamos problema mente-cuerpo tiene todos los visos de ser insoluble, si lo basamos en relaciones metafísico-formales tan generales como las de identidad, sobrevenida, realización o causalidad. Sus posibilidades de ser resuelto mermarán, si suponemos que nos enfrentamos, sin distinción, a nociones tan homogéneas como las de propiedad mental o propiedad física. De hecho, aquellos filósofos que plantean la cuestión en este grado de generalidad afirman que el problema es insoluble, es el caso de Colin McGinn (1991).

McGinn identifica el problema mente-cuerpo con el problema de la consciencia. Admite, como un hecho incontestable, que exista una propiedad natural que dé cuenta de la consciencia en términos naturalistas (es decir, la consciencia es un fenómeno físico), pero niega que podamos acceder alguna vez a ella. En consecuencia, no hay problema mente-cuerpo. Su conclusión se basa en que para llegar a comprender si la consciencia es una propiedad natural de los sistemas físicos, tendría que existir una forma de «homogeneidad» entre la manera en que comprendemos el fenómeno de la consciencia (introspectivamente, a partir de nuestro propio acceso al fenómeno) y la manera en que lo comprendemos físicamente, y tal forma de homogeneidad no existe. Por tanto, el fenómeno nos resultará para siempre ininteligible, pese a que

su existencia natural sea tan poco misteriosa como lo pueda ser la manera en que el hígado segrega la bilis.

Según nuestro autor, cualquier solución adecuada al problema de la consciencia consistirá en identificar una propiedad cerebral P y, a su juicio, ésta debería ser una propiedad de la que tuviésemos ya, al menos en un cierto sentido, una concepción clara. McGinn subraya, sin embargo, que incluso en el caso de que lográsemos llevar a cabo un análisis completo acerca de la naturaleza de P, seguiríamos sin comprender su conexión con las propiedades conscientes de nuestra experiencia. Ni la introspección, ni el conocimiento físico del cerebro pueden captar el vínculo entre P y lo que llamamos la consciencia.

Si formulamos el problema en estos términos, podríamos sentirnos inclinados a darle la razón a McGinn, cuyo punto de vista se acerca mucho a lo que, en ocasiones, se ha denominado la teoría «del doble aspecto»: los fenómenos mentales son fenómenos físicos, pero el aspecto bajo el cual nosotros los aprehendemos nos impedirá siempre concebirlos como físicos. En efecto, así planteada, la cuestión no parece tener solución. Pero no logramos comprender la relación entre la experiencia consciente y las propiedades cerebrales, precisamente porque la planteamos en este grado de generalidad. El dilema sugerido por McGinn es insoluble, si nuestra única elección se basa en un conocimiento introspectivo de la consciencia o en el conocimiento de una propiedad física. La neuropsicología y la psicopatología abren, no obstante, una tercera posibilidad al estudiar –nuevamente mediante un método controlado de variaciones– las relaciones entre las diversas lesiones y los desórdenes de la consciencia. Estos estudios no revelan cuál es la propiedad responsable de la experiencia consciente, sino una red de relaciones funcionales propias de la cognición normal. Aquí también, puesto que no es evidente que tengamos una propiedad mental homogénea (la consciencia), puede que la razón por la cual esta propiedad nos parezca físicamente indescifrable se deba a esta ausencia de especificidad.

Cuanto llevamos dicho parece sugerir que cualquier afirmación relativa a las propiedades mentales cuanto más general sea y más se refiera a entidades homogéneas, más posibilidades tiene de seguir siendo *a priori*, es decir, más posibilidades tiene de ser, a lo sumo, verdadera aunque inverificable. Sin embargo, es posible extraer una conclusión bien distinta de estos análisis. En efecto, los problemas fundamentales de la filosofía de la mente, como el de la relación mente-cuerpo, o el de la naturaleza de la intencionalidad, no tienen por qué ser considerados problemas, por definición, insolubles, que dan lugar a proposiciones inverificables. Al contrario, cabe sostener que las teorías construidas por los filósofos pueden ser testadas y evaluadas, en particular desde el punto de vista de su coherencia y de su plausibilidad empírica. Pero, puesto que no puedan serlo directamente, habría que intentar determinar un nivel intermedio de resolución de problemas y de interacción entre el nivel de descripción «filo-

sófico» y el nivel de descripción neurofisiológico (neuropsicológico, según el caso). Lo que sigue pretende mostrar cómo es posible hacerlo en relación con ciertos análisis acerca de la unidad de la consciencia.

III. LA UNIDAD DE LA CONSCIENCIA

En la *Analítica trascendental*, Kant sostiene que la unidad de la consciencia para un sujeto —entendiendo por tal la posibilidad de que diferentes contenidos de consciencia estén asociados a un único centro— requiere que el sujeto sea capaz de experimentar un mundo objetivo. Es este un argumento que normalmente se suele calificar de «trascendental»: parte de un hecho de experiencia (la consciencia está unificada) y se remonta a su condición de posibilidad (que exista un mundo objetivo para un sujeto).

Bernard Williams (1978) recoge otro argumento que conduce a la misma conclusión, la objeción del físico y aforista del siglo XVIII, Georg Lichtenberg al *cogito* cartesiano: ¿por qué tendríamos que afirmar mejor «Yo pienso» para referirnos a un sujeto de nuestros pensamientos en lugar de «se piensa» que no contiene ninguna referencia a tal sujeto? Williams nos aclara que la objeción de Lichtenberg no es correcta, ya que si observamos que «se piensa esto: P» y a continuación que «se piensa esto: Q», de ello se debería poder seguir «se piensa esto: P y Q». Sin embargo, este pensamiento conjuntivo podría no producirse, ya que los dos pensamientos simples están separados, pudiendo así surgir en dos consciencias distintas. Pero, si este tipo de pensamientos conjuntivos es simplemente posible, los dos pensamientos simples también deberían poder surgir en la misma consciencia⁴. En otras palabras, si es posible una unidad de la consciencia, debe de haber algo así como un principio de composicionalidad: *si se piensa P en i y si se piensa Q en i, entonces se piensa P y Q en i*, siendo *i* un índice que denota cualquier lugar en donde se produzca tal pensamiento.

Supongamos ahora que P y Q son contradictorios. No habría problemas en admitir que P y Q sean pensados simultáneamente, si son pensados en dos consciencias o en dos índices distintos. Pero en ese caso, no habría forma de determinar si nos estamos refiriendo a un único objeto o a dos, es decir, si hemos atribuido a un mismo objeto una propiedad y su contradictoria, o dos propiedades contradictorias a dos objetos diferentes. En otras palabras, para que pueda haber consciencia de dos objetos distintos, es necesario que el principio de composicionalidad antes mencionado sea posible o, lo que es lo mis-

4 cf. B. Williams, 1996, pp. 120-121.

mo, es necesario que puedan existir episodios de co-consciencia o que la consciencia tenga una unidad⁵.

Un neurocientífico que hubiera seguido este razonamiento hasta aquí probablemente lo habría identificado como un tipo de razonamiento, totalmente *a priori*, acerca de la consciencia, que no debería tener ninguna incidencia en la investigación en neurociencias. Pero ocurre que el neurocientífico parece enfrentarse a casos en los que el principio de composicionalidad antes descrito resulta infringido y, por tanto, casos en los que la unidad de la consciencia resulta problemática. Algunos trabajos sobre las comisurotomías⁶, como los de Justine Sergent (1990), parecen ejemplificar esta posibilidad.

A fin de comprobar la capacidad que tienen los sujetos comisurotomizados de unir y combinar información, esta neuróloga canadiense ha ideado un experimento en el que a los enfermos se les presenta –simultáneamente, uno hacia cada hemisferio– distintos números (por ejemplo, 6 y 7) y, a continuación, se les pide que los compare. El experimento se desarrolla en tres fases. En un primer momento, previa presentación de los dos números, se pregunta a los sujetos si éstos son iguales o diferentes. Los sujetos no responden correctamente a la pregunta o lo hacen al azar. En un segundo momento, se les pregunta si uno de los dos números es mayor que el otro. En este caso, la tarea les resulta fácil y responden correctamente. Y en un tercer momento, los enfermos deben determinar si los dos números son iguales o si uno es mayor que el otro. Los sujetos ofrecen una respuesta correcta sin vacilar. La cuestión que se plantea entonces es la siguiente: ¿cómo es posible que los enfermos comisurotomizados no sean capaces de responder de forma aislada a cuestiones relativas a la identidad de dos números, y sin embargo sean capaces de responder a preguntas que se refieren a un tiempo a su igualdad o a su tamaño? De los resultados obtenidos parece desprenderse que la información concerniente a la cantidad comparada es compartida por los dos hemisferios, pero no ocurre lo mismo con aquella relativa a la identidad. Semejantes casos parecen

5 S. Hurley retoma este razonamiento de Williams en su trabajo «Unity and objectivity?» en C. Peacocke, 1994.

6 La comisurotomía o seccionamiento del cuerpo calloso (conjunto de fibras nerviosas que une los dos hemisferios cerebrales) es una operación que se empezó a practicar, en los años 50, a sujetos humanos afectados de epilepsia. Tenía como finalidad corregir los síntomas de la enfermedad, en una época en que aún no se contaba con quimioterapias adecuadas. La intervención conseguía evitar una propagación interhemisférica de la crisis, a la par que preservaba las capacidades de autocontrol del enfermo. Roger Sperry puso de manifiesto que la bisección de los hemisferios cerebrales planteaba la posibilidad de una división de las actividades mentales del enfermo en dos flujos independientes de consciencia. Pese a ello, la unidad y la cohesión de su conducta no se hallaba comprometida. *cf.* Sperry, 1968, p. 724.

ilustrar de un modo directo situaciones de no-composicionalidad o de no-unidad de la consciencia.

Susan Hurley (1994), que desde el ámbito de la filosofía examina estos casos, señala que rebaten la tesis de una unidad fuerte de la consciencia, pero no la de una unidad *débil*, lo cual explicaría por qué los sujetos, aun no siendo capaces de determinar la identidad de los números que se les presenta, sin embargo sí logran distinguir propiedades parcialmente constitutivas de su identidad. De ello se deduce que sigue siendo cierto –tal como sostiene el argumento kantiano– que esta unidad débil de la consciencia requiere la percepción de propiedades objetivas de los objetos.

Este tipo de argumentación resulta de interés en la medida en que se ha partido de una caracterización puramente *a priori* de la unidad de la consciencia, no para tratar de rebatirla o confirmarla directamente con los datos neurológicos experimentales, sino más bien para modificarla. Con todo, conviene ser prudentes a la hora de valorar esta apreciación positiva. La propiedad de la consciencia de la que nos hemos ocupado aquí es la propiedad de una experiencia directa, relatada en juicios. Pero no resulta evidente que casos como los estudiados por Sergent se refieran realmente a este tipo de consciencia. Hay que tener en cuenta que algunas de las respuestas ofrecidas por los sujetos comisurotomizados no se basan en estados conscientes (en algunos casos los enfermos adivinan la respuesta correcta) y los fracasos de la composicionalidad deben darse en estados que se producen en un mismo momento. Tal como señala Anthony Marcel, para que se pueda hablar de una unidad de la consciencia, fuerte o débil, es necesario que estén presentes unas condiciones experimentales bastante específicas⁷. Por tanto, estas comparaciones entre proposiciones filosóficas *a priori* y observaciones experimentales deben efectuarse con toda clase de precauciones.

Pese a las dificultades, el razonamiento que precede puede ilustrar un tipo de investigación fecundo, capaz de propiciar interacciones entre filosofía y neurociencias. Se trata de partir de argumentos *a priori* con respecto de una noción. Un argumento *a priori* es un argumento que consiste en proponer la condición necesaria de un cierto fenómeno mental (por ejemplo, en nuestro caso, la composicionalidad de los estados co-conscientes). Esta condición, por su parte, implicará ciertas tesis contrafácticas acerca de lo que debería ocurrir si tal condición no estuviese presente (en nuestro caso, implicaría que la consciencia debería estar dividida). Habitualmente, los filósofos no van más allá, y estas tesis contrafácticas nunca son verificadas, si no es por medio de las lla-

⁷ cf. A. Marcel, «What is relevant to the unity of consciousness?», en C. Peacocke (ed.), 1994, p. 81.

madas «experiencias de pensamiento» que se basan, como su propio nombre indica, en nuestras intuiciones acerca de lo que es posible o concebible en semejante situación. Éste es el tipo de razonamiento que se sigue, por ejemplo, en relación con la noción de identidad personal. Se plantean diversos tipos de situaciones imaginarias en las cuales esta identidad se rompería⁸. Pero existe la posibilidad de perfilar—especialmente si tenemos en cuenta las investigaciones realizadas en neuropsicología y neuropatología— estructuras de disociación y contrastes, entre cognición normal y cognición anormal, que pueden ser considerados también como verificaciones posibles de dichas tesis contrafácticas. En algunos casos, los argumentos *a priori* serán verificados; en otros, serán revisados. Por tanto, no hay razones para pensar que estos argumentos hayan de ser condenados a seguir siendo puramente *a priori*. Cabría ilustrar esta perspectiva de estudio con otro problema, enormemente discutido en filosofía de la mente. Me refiero al de la naturaleza de la representación o, lo que algunos filósofos de la mente han denominado, el problema del «contenido» de los estados mentales y más en concreto a uno de sus aspectos más destacados.

IV. NATURALEZA DE LA REPRESENTACIÓN MENTAL

Una de las hipótesis más atractivas de la investigación en ciencias cognitivas es aquella que fundamenta la estructuración de las representaciones en un lenguaje o código interno, lo que Fodor ha denominado el «lenguaje del pensamiento». Es característico de toda una corriente filosófica de la mente discutir esta hipótesis, en términos puramente *a priori*, es decir, mediante argumentos que se supone nos deben mostrar por qué ha de existir un lenguaje del pensamiento. La siguiente argumentación propuesta por Fodor (1987) estaría en esta línea:

- a) Los rasgos propios o propiedades características del pensamiento son su *sistematicidad* (si podemos pensar que Juan ama a María, debemos poder pensar que María ama a Juan), su *productividad* (si podemos pensar que Juan ama a María, debemos poder pensar también que Raúl ama a María, o que hay personas que se aman, etc.) y sus *poderes causales* (si yo pienso que si nieva, hace frío, y si llego a pensar que nieva, en general pensaré que hace frío);
- b) Estas tres propiedades quedan perfectamente explicadas si suponemos que el pensamiento está codificado en un lenguaje. Concretamente, la

⁸ *cf.*, por ejemplo, D. Parfit, 1984.

tercera resulta perfectamente explicable si suponemos que las estructuras físicas de los símbolos del lenguaje mental tienen un poder causal sobre otras estructuras físicas.

Existe, sin embargo, otra opción posible, ya que no todas las representaciones son necesariamente de tipo lingüístico. Tal como señala Pascual Martínez Freire en su trabajo «Representación e interpretación» (1993), podemos afirmar que R es una representación de un objeto O para un intérprete I en la medida en que I toma cuenta de O en virtud de la manifestación de R. Teniendo en cuenta esta caracterización general de la representación, los mapas, las imágenes, los diagramas también pueden ser considerados representaciones. Desde luego, proporcionan información, incluso más información que una representación lingüística y no poseen articulaciones evaluables semánticamente de manera discreta (cada parte de un mapa contribuye al contenido representacional del conjunto). Pero los mapas, las imágenes, etc. también tienen propiedades de sistematicidad, productividad y causalidad. Podemos reestructurar un mapa para que represente, por ejemplo, Madrid al sur de Málaga en lugar de a la inversa; un cartógrafo puede representar tantas situaciones nuevas como alguien que utilice frases, y los mapas son, al igual que los símbolos de un lenguaje, estructuras físicas que tienen poderes causales.

En consecuencia, la idea de que un gran número de representaciones mentales pudieran estar más próximas de los mapas que de las frases de un lenguaje no está menos justificada *a priori* que la hipótesis de un lenguaje mental. Tal idea se ajusta tanto a las propiedades de estados intencionales, en las que han insistido mucho algunos filósofos⁹, como a las numerosas investigaciones en neurociencias y psicología cognitivas¹⁰.

Sin embargo, el hecho de que un gran número de trabajos en ciencias cognitivas pueda apoyarse en esta hipótesis aún no constituye un ejemplo de argumento *transcendental*. Para que se diera un argumento de este tipo, sería necesario derivar de toda representación espacial una condición necesaria, para verificarla luego con los datos experimentales.

Es de todos conocida la condición transcendental o *a priori* del espacio tal como la expuso Kant en la *Estética transcendental*: toda representación de un universo objetivo es necesariamente espacial. Para comprender el planteamiento kantiano hemos de distinguir dos tipos de condiciones que Kant denomina,

9 Como por ejemplo, el *holismo*: la manera en que yo crea que la nieve es blanca no corresponde necesariamente a una creencia que tenga ese contenido, sino a muchas otras que están unidas a ella.

10 Recuérdense los «mapas cognitivos» en animales y humanos de E. Tolman y J. O'Keefe; los «modelos mentales» de P. Johnson-Laird en psicología del razonamiento, etc.

respectivamente, condiciones empíricas y condiciones *a priori*. Consideremos el caso de la visión. El que podamos ver algo depende de un sinnúmero de condiciones. Depende, en primer lugar, de nuestra agudeza visual o del tamaño del objeto. Estas condiciones son particulares (afectan a la visión de un determinado objeto o de un individuo en particular) y meramente fácticas (pueden ser alteradas). Este tipo de condiciones se denomina condiciones empíricas. Pero la visión depende también –y esto es lo verdaderamente importante para Kant– de dos condiciones absolutamente generales y necesarias, el espacio y el tiempo. No se puede ver algo, sin verlo en un lugar del espacio y en un momento del tiempo. Este ejemplo puede generalizarse: espacio y tiempo son condiciones generales y necesarias –trascendentales– no sólo de la visión sino de la sensibilidad.

Para ser más concretos, Kant considera que el espacio es «una forma *a priori* de la sensibilidad» o del conocimiento sensible. Es decir, el espacio no es una impresión sensible particular como los distintos colores, sonidos, etc., sino la forma o el modo como percibimos todas las impresiones particulares: los colores, los sonidos, etc. son percibidos en el espacio. Es una forma «*a priori*», en el sentido de que es previa a la experiencia. No proviene de la experiencia, sino que condiciona a ésta. Es una condición que pertenece a la estructura del sujeto (todo sujeto que percibe por medio de los sentidos, percibe necesariamente en un lugar del espacio) El espacio no procede de la experiencia, sino que la precede, como condición para que ésta sea posible. Por tanto, las representaciones mentales que reconstruyen parte de la realidad se producen necesariamente en el espacio.

John O'Keefe, en un trabajo titulado *Cognitive maps, time and causality* (1994) dedicado a este problema kantiano, sostiene que las propiedades de los mapas cognitivos en las ratas¹¹ son tales que estas representaciones espaciales reproducen parcelas más o menos amplias del universo físico y son anteriores a la noción de objeto físico e independientes de él. O'Keefe indica que los mapas cognitivos son objetivos en la medida en que generan representaciones del entorno independientes del organismo y de sus acciones.

Probablemente cabría objetar que estas discusiones pueden conducir a los neurocientíficos cognitivos a pronunciarse acerca de nociones ontológicas que, de hecho, no tienen mucho que ver con su trabajo real. Pero este tipo de análisis no es ni mucho menos nuevo en neurociencias ni en psicología cognitivas.

¹¹ J. O'Keefe junto a L. Nadel, en *The Hippocampus as a cognitive map* (1978), ya puso de manifiesto –basándose en el estudio del hipocampo y sus lesiones en ratas– la capacidad que tienen los organismos de representar el espacio en mapas cognitivos. Véase al respecto los análisis de P. Martínez Freire en «Epistemología con sujetos cognitivos», 1998.

Los trabajos de Nancy Soja, Susan Carey e Elizabeth Spelke acerca de la percepción de objetos en los niños están directamente enfocados a ciertos argumentos trascendentales –como los de Quine (1960), para quien ciertas categorías (objeto o sustancia) emergen a partir de una conceptualización de los objetos en términos de masa y no de individuos discretos– y apuntan a su refutación (Spelke, 1990; Soja, Carey y Spelke 1991).

V. FILOSOFÍA Y APROXIMACIÓN NEUROCIENTÍFICA DE LA MENTE

El estudio de estos casos nos ha permitido poner en evidencia una posible trama de argumentación situada en el punto de encuentro entre filosofía y neurociencias. Se trataría básicamente de plantear primero ciertas propiedades necesarias o ciertas condiciones de posibilidad para el ejercicio de ciertas actividades cognitivas, de ciertos conceptos o de ciertas representaciones mentales, con objeto de determinar después en qué medida los trabajos de las neurociencias cognitivas permiten confirmarlos, rebatirlos o modificarlos. Ésta podría ser la táctica a seguir por parte de la neurofilosofía, perfilándose de este modo un ámbito de estudio intermedio entre la especulación filosófica (siendo un buen ejemplo de ello, la discusión formal del problema mente-cuerpo o de las relaciones de sobreveniencia) y las explicaciones científicas de las neurociencias cognitivas. La neurofilosofía determinaría así las condiciones generales que deben cumplir los procesos y estados mentales –las representaciones, los conceptos o las experiencias– para que puedan llevar a cabo tareas cognitivas.

Quizá a la luz de las afirmaciones precedentes puede que, desde el ámbito de la filosofía, algún lector se pregunte si la neurofilosofía así concebida no es sino un campo imaginario. Es muy posible que entre los defensores ortodoxos de Kant la respuesta sea afirmativa, ya que probablemente rechazarían cualquier injerencia de la ciencia en la filosofía y podrían identificar el tipo de investigación aquí sugerida precisamente con una psicología de las facultades a la que alude Kant en la *primera* edición de la *Crítica de la razón pura*. Sin embargo, la respuesta debe ser negativa: no sólo no es un campo imaginario, sino que corresponde precisamente a un conjunto de trabajos contemporáneos que pretenden «naturalizar» el trascendental kantiano, como he tratado de sugerir aquí con algunos ejemplos.

Por su parte, los neurocientíficos cognitivos quizá podrían objetar que nos hallamos ante una temática que a todas luces resulta interesante para los filósofos –que siempre tienen cuentas que saldar con Kant–, pero que poco les aporta a ellos ¿Acaso lo que se ha discutido no es una simple variante del método hipotético-deductivo utilizado en ciencias experimentales: se propone una hipótesis, se deducen posibles observaciones y se la verifica? La respuesta es nuevamente

negativa. Las hipótesis «trascendentales» o *a priori* no son precisamente hipótesis empíricas sino hipótesis que descansan únicamente en la imaginación conceptual, siendo ésta posteriormente confrontada con los datos experimentales.

No ha sido nuestra intención afirmar que este tipo de interacción entre filosofía y neurociencias deba ser considerado el único posible y fructífero. Tampoco se ha querido excluir la manera más común de trabajar, para el filósofo, consistente en sumergirse en las actuales teorías de las neurociencias e intentar extraer de sus trabajos las teorías de la mente más plausibles. Tan sólo se ha pretendido sugerir una perspectiva también fecunda de afrontar las relaciones entre filosofía y neurociencias, sin duda compatible con aquella otra más tradicional.

El desarrollo de una neurofilosofía así concebida, como punto de encuentro entre filosofía y neurociencias, vuelve a replantear la discusión acerca de las relaciones y fronteras entre disciplinas. Este campo emergente puede ser contemplado por parte de algunos observadores como el triunfo del reduccionismo neurofisiológico propio de las neurociencias y la consiguiente absorción por parte de éstas de la filosofía, tal como propugnara Patricia Churchland. Sin embargo, la reducción y posterior absorción no son las únicas maneras de entender las relaciones entre ambas disciplinas. También es posible la interacción y la reflexión conjunta, como hemos visto. En efecto, las fronteras entre disciplinas son porosas y están en permanente evolución, de tal manera que los pensadores situados en ambos lados de estas fronteras saben lo que les separa y lo que, en general y no en contenido, permanece inamovible.

¿Cuál debe ser entonces, en la actualidad, el papel de la filosofía frente a una aproximación científica de la mente? Ante todo, hay que subrayar que al aceptar trabajar en contacto directo con los científicos, tanto en los problemas particulares como en aquellos relativos al marco teórico que se debe adoptar, el filósofo abandona esa posición de predominio, o de separación, que fue la suya en otras épocas o que sigue teniendo en otros ámbitos de la filosofía. Pero al mismo tiempo supone un desmentido a la idea de que se trata de un juego sin pérdidas ni ganancias en el que el terreno ganado por la ciencia, en el proceso de especificación de problemas, ha sido perdido por la filosofía. En el universo de las ciencias contemporáneas, la filosofía no parece correr más peligro de debilitación que el Estado en las sociedades avanzadas. Al contrario, su tarea se ha hecho más laboriosa y compleja. Es más, me atrevería a decir que hoy en día resulta más difícil ser filósofo de la física, por ejemplo, que en tiempos de Galileo. Y no sólo porque las teorías físicas sean más profundas y complejas hoy que entonces, sino porque los problemas ontológicos, lejos de desaparecer, se han desplazado. Son las soluciones fáciles las que se desvanecen.

En el ámbito de la neurofilosofía, la situación teórica del filósofo es quizá más difícil si cabe. No porque las teorías en este ámbito tengan una profundidad comparable a las de la física, sino porque el filósofo tiene en ellas un papel más directamente activo, precisamente en razón de la juventud de este dominio. Esto le obliga por una parte a trabajar más, por otra a hacer un esfuerzo mayor para conservar su independencia crítica. Y esta sigue siendo indispensable. Podríamos decir que filósofo dedicado a la neurofilosofía se halla atrapado entre dos actitudes. La primera, tradicional consiste en permanecer alerta y no dar prueba de credulidad. La segunda, característica en los inicios de una disciplina, consiste en llevar las hipótesis propuestas tan lejos como sea posible, sin detenerse ante las objeciones aparentes, ya tomen la forma de argumentos con sentido, de conflictos con las concepciones científicas o filosóficas del momento, o de limitaciones evidentes con respecto a la extensión de los fenómenos estudiados, o incluso de contradicciones internas. Aquí radican sus dificultades. Por tanto, y pese a ellas, el neurofilósofo no sólo debe limitarse a seguir los resultados obtenidos o simplemente sugeridos por las neurociencias. Tiene el derecho también y, sobre todo, los medios para mezclarse en la investigación misma. En este punto de encuentro que es la neurofilosofía, a la filosofía (no toda ella desde luego, sino una parte no desdeñable, aquella que forma parte de las ciencias cognitivas) no le correspondería pues una posición de segundo plano, propia del comentarista o del historiador, sino de verdadera colaboración, interacción y reflexión conjunta.

REFERENCIAS BIBLIOGRÁFICAS

- DAVIDSON, D. 1980: *Essays on Actions and Events*. Oxford: Oxford University Press.
- DEHAENE, S. (ed.) 1997: *Le cerveau en action. Imagerie cérébrale fonctionnelle*. Paris: PUF.
- DEHAENE, S. 1997: «La méthode de soustraction» en S. Dehaene (ed.), *Le cerveau en action*, Paris: PUF, pp. 97-123.
- FODOR, J. 1987: *Psicosemántica: el problema del significado en la filosofía de la mente*, tr. O. González- Castán. Madrid: Tecnos, 1994.
- GOLDMAN, A. (ed.) 1993: *Readings in philosophy and cognitive science*, Cambridge, MA: MIT Press.
- HURLEY, S. 1994: «Unity and objectivity?» en C. Peacocke (ed.), *Objectivity, simulation and the unity of consciousness*. Oxford: Oxford University Press, (pp. 49-78).
- KANT, I. 1781/1787: *La crítica de la razón pura*, tr. Pedro Ribas. Madrid: Alfaguara, 1984.
- KIM, J. 1993: *Supervenience and mind: selected philosophical essays*. Cambridge: Cambridge University Press.
- LEPORE, E. y MCLAUGHLIN, B.(eds.) 1985: *Actions and Events. Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell.

- MARCEL, A. 1994: «What is relevant to the unity of consciousness?» en C. Peacocke (ed.), *Objectivity, simulation and the unity of consciousness*. Oxford: Oxford University Press, pp. 79-88.
- MARTINEZ FREIRE, P. 1993: «Representación e interpretación», *Themata*, nº 11, pp. 37-46.
- 1998: «Epistemología con sujetos cognitivos» en P. Martínez Freire (ed.), *Filosofía actual de la ciencia*, suplemento 3 de *Contrastes*, pp. 95-118.
- MCGINN, C. 1991: *The problem of consciousness*. Oxford: Basil Blackwell.
- O'KEEFE, J. y NADEL, L. 1978: *The hippocampus as a cognitive map*. Oxford: Oxford University Press.
- O'KEEFE, J. 1994: «Cognitive maps, time and causality» en C. Peacocke (ed.), *Objectivity, simulation and the unity of consciousness*. Oxford: Oxford University Press, pp. 35-48.
- PARFIT, D. 1984: *Reasons and Persons*. Oxford: Oxford University Press.
- PEACOCKE, C. (ed.) 1994: *Objectivity, simulation and the unity of consciousness: current issues in the philosophy of mind*. Oxford: Oxford University Press.
- QUINE, W. 1960: *Word and Object*. Cambridge, MA: MIT Press.
- SERGENT, J. 1990: «Furtive incursions into bicameral minds», *Brain*, nº 113, pp. 537-568.
- SPELKE, E. 1990: «Origins of visual knowledge» en D.N. Osherson, *Visual cognition and action. A invitation to cognitive science*, vol. 2, Cambridge, MA: MIT Press (reimpreso «Object perception» en A. Goldman, 1993, pp. 447-460).
- SOJA, N., CAREY, S. y SPELKE, E. 1991: «Ontological categories guide young children's inductions of word meaning», *Cognition*, nº 38, pp. 179-211 (reimpreso en A. Goldman, 1993, pp. 461-480).
- SPERRY, R. 1968: «Hemisphere deconnection and unity in conscious awareness», *American Psychologist*, nº 23, pp. 723-753.
- VAN GELDER, T. 1998: «Monism, dualism, pluralism», *Mind and Language*, nº 13, pp. 156-175.
- WILLIAMS, B. 1978: *Descartes: el proyecto de una investigación pura*. Madrid: Cátedra, 1996.

Alicia Rodríguez Serón es profesora asociada de Filosofía de la mente en la Universidad de Málaga. Es autora de *El estudio de la memoria en Bergson: antecedentes e implicaciones filosóficas* (Universidad de Málaga, 1998) y «Monismo anómalo, irreductibilidad y ciencias cognitivas», *Contrastes*, 5 (2000), pp. 161-178.

Dirección Postal: Departamento de Filosofía, Campus de Teatinos, Universidad de Málaga, Facultad de Filosofía y Letras, E-29071, Málaga.

E-mail: ars@uma.es