

Monismo anómalo, irreductibilidad y ciencias cognitivas

ALICIA RODRÍGUEZ SERÓN
Universidad de Málaga

RESUMEN

El pensamiento de Donald Davidson ocupa un lugar central en la filosofía actual. Su obra ha ejercido una gran influencia y ha suscitado numerosas discusiones que han interesado campos tan diversos como la filosofía de la mente, la psicología, la semántica, la lógica o las ciencias sociales. El presente trabajo pretende poner de manifiesto cómo lo que más seduce de la postura de Davidson es también lo que constituye la raíz de sus dificultades: el intento de reconciliación de un fisicalismo monista y un antireduccionismo.

PALABRAS CLAVE

DAVIDSON-MONISMO ANÓMALO-IRREDUCTIBILIDAD-CIENCIAS COGNITIVAS

ABSTRACT

Donald Davidson's thought occupies a central position within the current philosophy. His work exerted great influence and has provoked a number of discussions that have affected a wide range of fields such as philosophy of mind, psychology, semantics, logic or social sciences. This paper argues the fact that, what attracts most about Donald Davidson's view, is also what constitutes the root of its difficulties: the attempt of reconciliation between a monist physicalism and an antireductionism.

KEYWORDS

DAVIDSON-ANOMALOUS MONISM-IRREDUCTIBILITY-COGNITIVE SCIENCES

HOY MÁS QUE NUNCA, en filosofía, nos encontramos enfrentados a dos imágenes opuestas acerca de la mente que nos resultan ya familiares. Por un lado, la imagen naturalista, fisicalista y reduccionista que nos hace concebir la mente humana como una parte más de la naturaleza, susceptible de recibir una explicación causal semejante a la que reciben los fenómenos estudiados por las

ciencias naturales. Según esta primera imagen, lo que denominamos nuestra mente no sería más que un conjunto de entidades o de hechos físicos y materiales, o al menos estaría vinculada a ellos. La segunda, es la imagen intuitiva que tenemos de nosotros mismos en tanto que agentes libres, racionales y autónomos, difíciles de reducir a determinismos causales y naturales.

Son numerosas las versiones que pueden plantearse de este antagonismo: materia y espíritu (cerebro y mente), naturaleza y libertad, objetivo y subjetivo, ciencias naturales y ciencias sociales y humanas, positivismo y hermenéutica, por sólo citar algunas. Incluso el actual conflicto surgido entre filósofos partidarios de una «naturalización» de la intencionalidad y de lo mental (especialmente a partir de las ciencias cognitivas) y filósofos que rechazan tal naturalización puede ser considerado una muestra representativa de esta oposición familiar.

Semejante antagonismo parece ser irreductible de tal forma que adoptar una de las imágenes en detrimento de la otra supondría adentrarse en un evidente callejón sin salida. Por un lado, las concepciones naturalistas y reduccionistas de la mente sufren y se resienten por el hecho de que nunca se haya logrado consolidar ningún programa reduccionista, de modo que toda reducción de lo mental a lo físico parece conducirnos a la idea de que la mente y nuestra imagen intuitiva de nosotros mismos debe ser lisa y llanamente abandonada a favor de una descripción puramente fisicalista (materialismo eliminativo). Esto, a su vez, no puede por menos de suscitar la reacción –siempre molesta para los fisicalistas– de que algo se puede perder en el camino de la reducción. Por otro, la idea de que la mente simplemente estaría ahí, completamente fuera del alcance de cualquier explicación científica, resulta igualmente sorprendente: ¿acaso somos únicamente seres materiales en un mundo natural? En este dilema nos encontramos actualmente a la hora de explicar la mente.

Pues bien, el atractivo de la postura filosófica de Donald Davidson reside en haber intentado formular una forma de escapar a este dilema que, desde luego, no está exenta de dificultades y puntos oscuros, pero cuyas consecuencias resultan de interés en la medida en que inciden, en cierto sentido, en algunos planteamientos de las ciencias cognitivas. Me centraré en su teoría de los acontecimientos y su aplicación a las relaciones entre lo mental y lo físico y en su teoría de la interpretación. Sigamos la argumentación de Davidson en sus líneas maestras.

I. TEORÍA DE LOS ACONTECIMIENTOS

Davidson parte de la idea de que normalmente solemos admitir la existencia de objetos materiales o físicos: mesas, perros, esmeraldas, personas, etc.. Estos

objetos tienen propiedades: ser de madera, tener pelo, ser verde, ser mortal, etc. Los objetos son individuos, materiales o físicos, pero las propiedades no siempre tienen un estatus claro: no sabemos si son independientes de los individuos que las poseen, si son individuos abstractos. Lo que sí parece evidente es que las propiedades se expresan mediante predicados o descripciones que se consideran verdaderos con respecto a tales objetos materiales. Esto quiere decir que no hay entidades abstractas generales como el verdor, aunque sí exista el verdor de una esmeralda o de la hierba.

En su trabajo «The Logical Form of Action Sentences» (1967), junto a los objetos materiales o físicos y sus propiedades, Davidson introduce una nueva categoría, los acontecimientos, insistiendo en que éstos deben ser considerados como entidades particulares del mismo tipo que los objetos materiales y físicos. Serían acontecimientos una explosión, asistir a un concierto, levantarse tarde, etc. Estos acontecimientos se caracterizan y distinguen por sus causas y efectos. En principio, asistir a un concierto no se distingue de desayunar tarde, si ambos hechos se producen uno a continuación del otro. Lo que los distingue son sus causas y efectos. El asistir a un concierto causa que me acueste tarde, el acostarme tarde causa que necesite dormir más, mi necesidad de dormir más causa que me levante tarde. En cada uno de estos casos, se ha producido un cambio. Los cambios son acontecimientos, sus límites espacio-temporales pueden ser más o menos precisos al igual que aquellos propios de los objetos materiales y físicos.

Describimos los acontecimientos, nos dice Davidson, de la misma manera que describimos a los individuos. Un mismo acontecimiento puede ser descrito de distintas maneras. Una explosión puede ser descrita como un ruido que procede de la calle, como una chispa que ha prendido una bombona de butano, como un atentado que se ha producido esta mañana, etc. Solemos describir los acontecimientos utilizando el lenguaje. Nuestras descripciones lingüísticas son la expresión de nuestras descripciones mentales; describimos el mundo de una determinada manera porque poseemos ciertos conceptos, ciertas creencias, ciertas percepciones, ciertos juicios. Pero tener creencias, percepciones, juicios, también son acontecimientos, aunque puede que su duración sea distinta. Son acontecimientos mentales. Mi creencia de que Málaga está en Andalucía, puede durar casi toda mi vida y ésta no es menos acontecimiento que aquella otra que pueda tener de que el 22 de abril el sol brilla en Málaga, que es menos duradera. Las creencias y las percepciones son acontecimientos mentales causados por las cosas del mundo. Pero también hay acontecimientos mentales que pueden causar acontecimientos del mundo: los deseos, intenciones, voliciones y acciones. Por ejemplo, mi intención de rascarme la mejilla derecha causa el movimiento de mi mano derecha hacia mi mejilla derecha y el rascamiento de mi mejilla derecha.

Así como describimos los sucesos físicos mediante ciertos predicados y descripciones físicas más o menos precisos, más o menos acertados, así describimos también los acontecimientos mentales. Nuestras descripciones más comunes suelen ser descripciones hechas, como acabo de señalar, en términos de creencias, deseos, intenciones y de todo ese conjunto de lo que se denomina «actitudes proposicionales». De la misma manera que describimos los acontecimientos físicos en función de sus causas y efectos, así también describimos los acontecimientos mentales. Así por ejemplo, decimos que *porque* José María creía que su coche estaba en el aparcamiento se dirigió a éste. Al menos en principio, señala Davidson, el vocabulario por el que normalmente expresamos los motivos de las acciones es un vocabulario causal.

Hay que notar además que dos acontecimientos individuales (*tokens*) pueden ejemplificar o instanciar un único y mismo tipo (*type*) de acontecimiento: un atentado en Madrid o en San Sebastián ejemplifican el mismo tipo general de acontecimiento, *atentado*.

Esta distinción entre acontecimientos y descripciones y entre *tokens* y *types* de acontecimientos, Davidson la aplica también a los acontecimientos mentales, tales como una sensación, o la ocurrencia de un pensamiento. Por ejemplo, mi creencia de que el 22 de abril de 2000 luce el sol en Málaga puede ser descrita de dos formas distintas: desde un punto de vista neurofisiológico, como un acontecimiento físico que ocurre en mi cerebro y, desde un punto de vista mental, como un acontecimiento mental expresado en términos intencionales que recogen una actitud proposicional.

Decir que estas dos descripciones pueden aplicarse al mismo acontecimiento o que estos dos *tipos* pueden ser instanciados por el mismo *token*, es lo mismo que decir que el acontecimiento mental *mi creencia de que el 22 de abril de 2000 luce el sol en Málaga es idéntico* al acontecimiento físico que yo describiera, por ejemplo, como la activación de ciertas neuronas en mi cerebro. Por consiguiente, los acontecimientos mentales (individuales, *tokens*) son acontecimientos físicos (*tokens*). Pero de ello no se sigue que las descripciones mentales y físicas o los tipos mentales y físicos que ejemplifican estos acontecimientos-tokens sean ellas mismas idénticas, ni que los primeros se reduzcan a los segundos. En otras palabras, Davidson defiende una teoría de la identidad de casos (*token-token*) de la mente y del cerebro, lo que a veces se denomina el materialismo de los individuos, pero no una teoría de la identidad de *tipos* mentales y *tipos* físicos usualmente denominada teoría de la identidad de tipos (*type-type*).

Davidson parte pues de la idea de que el mundo, lo que hay, está compuesto de individuos (principalmente físicos), de propiedades (principalmente propiedades concretas que se expresan mediante predicados) y de relaciones (principalmente espacio-temporales y de causalidad). Añade a esta ontología muy

austera una categoría más, la de acontecimiento, considerada en el mismo plano que la de los individuos concretos. Cabe preguntarse a continuación cómo entiende nuestro autor la relación entre las propiedades y los predicados físicos de los individuos físicos, por una parte, y las propiedades y predicados mentales de estos individuos, por otra.

II. MONISMO ONTOLÓGICO, NO MONISMO EPISTEMOLÓGICO

Según el dualismo ontológico, hay individuos físicos e individuos mentales (dualismo de sustancias) o propiedades mentales y propiedades físicas (dualismo de propiedades). Según el monismo o materialismo ontológico, sólo hay una categoría de seres, individuos o propiedades físicas, y los individuos (si los hay) o propiedades mentales (si las hay) son idénticas o reductibles a individuos o propiedades físicas. Según el esquema ontológico que acabamos de ver, la teoría davidsoniana sería una forma de monismo materialista. Entendiendo por tal la tesis según la cual el mundo está constituido únicamente por objetos, propiedades y acontecimientos que son verdaderos en el lenguaje de las ciencias físicas. Sin embargo, tal monismo materialista queda fuertemente matizado por un argumento central en Davidson. Las tres premisas¹ fundamentales en las que se basa su argumento son las siguientes:

- 1) Todos los acontecimientos mentales están causalmente unidos a acontecimientos físicos. Esto resulta bastante evidente para Davidson y ya lo hemos señalado: las creencias, intenciones y deseos causan acciones físicas. A la inversa, los acontecimientos físicos causan modificaciones de nuestras creencias, intenciones y deseos.
- 2) Si dos acontecimientos están unidos como causa y efecto, hay una ley estricta bajo la cual pueden estar subsumidos. Dicho de otra manera, es posible describir la causa y el efecto de dichos acontecimientos en forma de descripciones que instancien una ley estricta. Davidson entiende por ley estricta una ley que adopta la forma «todo X causa Y» en todos los casos, es decir, sin cláusulas *ceteris paribus*. Por ejemplo, «un río con meandros erosiona su orilla externa» no es una ley estricta, la ley de Coulomb sí². Hay quien afirma que quizá no haya leyes estrictas en física. Pero, incluso si no las hubiese, sostiene Davidson, eso no cambiaría nada. Lo importante es que las leyes estrictas sean las que carac-

¹ Cf. D. Davidson, «Thinking Causes» en J. Heil y A. Mele (eds.), *Mental Causation* (1995); «Donald Davidson» en S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*. Oxford: Blackwell, 1994, pp. 231-235, y «Mental Events» (1970), pp. 209-210, 213.

² Cf. J. Fodor, *Psicosemántica*, Tecnos: Madrid, 1994, p. 22.

tericen un sistema cerrado y que éstas sean tales que no puedan convertirse en falsas mediante cláusulas *ceteris paribus*. No tienen por qué ser necesariamente las leyes de la física, sino de cualquier ciencia que pretenda describir sistemas cerrados mediante leyes. Si la neurofisiología fuese una ciencia así, entonces 2) sería también verdadero. Davidson considera que este principio es desde luego discutible, pero ha sido ampliamente aceptado por Hume y por Kant, por ejemplo.

- 3) No hay leyes psicofísicas estrictas, es decir, no hay leyes estrictas que conecten un acontecimiento descrito como mental con un acontecimiento descrito como físico³. Esta tercera premisa supone un rechazo claro del conductismo según el cual los predicados mentales deben ser definidos en términos conductuales o físicos y también de la teoría de la identidad que establece una conexión estricta entre las propiedades mentales y las propiedades físicas.

El monismo que se obtiene aquí no es un monismo de propiedades: no afirma que las propiedades o las descripciones mentales sean idénticas a las propiedades o descripciones físicas. Es un monismo de individuos o más exactamente de acontecimientos. Sostiene que los acontecimientos particulares mentales son idénticos a acontecimientos físicos, pudiendo instanciar o ejemplificar propiedades o descripciones físicas. En otras palabras, únicamente hay un tipo de cosas, de individuos y de acontecimientos físicos, pero las descripciones, mentales o físicas, que podemos ofrecer de estos individuos no son idénticas ni reductibles unas a otras. Esta imposibilidad de reducción, apostilla Davidson, se debe al hecho de que no podemos ofrecer, para las descripciones mentales, leyes del mismo tipo que las que podemos ofrecer para las descripciones físicas o neurofisiológicas. De ahí el apelativo que recibe esta tesis de monismo anómalo. Hay monismo ontológico, pero no monismo explicativo o epistemológico, porque las leyes (si las hay) que rigen el universo de los conceptos mentales no se pueden reducir a leyes que rigen el universo de los conceptos físicos.

¿Qué consecuencias se desprenden de esta tesis? Sin duda varias. Detengámonos en una que puede incidir en las ciencias cognitivas. Una de las tesis

³ En tanto que son acontecimientos físicos, los acontecimientos mentales quedan subsumidos bajo leyes físicas y entran en relaciones causales nomológicas. Pero en tanto que acontecimientos mentales, no quedan subsumidos por tales leyes. Davidson piensa que no existen leyes para los tipos mentales y psicológicos, ni leyes psico-físicas que relacionen tipos mentales y tipos físicos de acontecimientos. Especialmente por esta razón, piensa que la psicología no puede ser una ciencia y que los conceptos mentales e intencionales descansan en principios completamente diferentes de los que rigen el universo físico. Cf. «Philosophy of Psychology», p. 241.

básicas que sostiene este conjunto de ciencias es que el sujeto de conocimiento es un sistema de procesamiento de la información, concretamente, un sistema de símbolos físico compuesto de una memoria, un control, un conjunto de operadores, una entrada y una salida. Los procesos mentales son caracterizados como procesadores de la información y descritos en términos funcionales como funciones mediadoras entre entradas sensoriales y salidas motoras, que pueden ser realizadas en distintos tipos de soportes⁴. En sentido restringido, los soportes han de ser ante todo de carácter físico; en sentido amplio, los soportes pueden ser de distinta índole.⁵ Así, teniendo en cuenta el primer sentido, a menudo, se suele afirmar que las ciencias cognitivas –que son ciencias que versan sobre el tratamiento de la información mental, por tanto, ciencias que se refieren a los acontecimientos mentales– en cierto sentido, están abocadas, desde el punto de vista ontológico, al materialismo, es decir, sólo postulan entidades físicas: símbolos, neuronas, cantidades de información descritas en términos físicos, etc. Con frecuencia a este monismo materialista ontológico, se le asocia un monismo epistemológico: las ciencias cognitivas no deben postular otras explicaciones de los acontecimientos mentales que no sean físicas, neurológicas o de cualquier otra ciencia que pueda ser reducida en última instancia a la física, al menos idealmente.

De acuerdo con el monismo anómalo planteado por Davidson, el postulado ontológico es correcto, pero el postulado epistemológico no. ¿Entraría esto en contradicción con las ciencias cognitivas? Todo depende del grado de compromiso que se tenga con respecto al monismo epistemológico; todo depende del grado de reducción que se asuma. Es de suponer que el monismo anómalo no será una teoría aceptable para aquellos teóricos de las ciencias cognitivas que asuman un funcionalismo en sentido restringido (basado en la hipótesis del sistema de símbolos físico), hagan hincapié en el postulado epistemológico y consideren que las descripciones mentales deben ser reducidas a descripciones físicas. Será aceptable para aquellos que insistan en el postulado ontológico: sólo hay acontecimientos físicos, la relación causal entre lo mental y lo físico se produce en el nivel físico. Finalmente, será una teoría indiferente para los cognitivistas que no insistan en ninguno de los dos postulados, y asuman un funcionalismo, en sentido amplio, según el cual los acontecimientos mentales –caracterizados según la función que desempeñan– son independientes del soporte físico y probablemente de la descripción que se haga de ellos.

⁴ Para el desarrollo de las tesis básicas de las ciencias cognitivas, véase P. Martínez Freire «El desafío de las emociones a las ciencias cognitivas», *Thémata*, 25 (2000), pp. 56-57, así como *La nueva filosofía de la mente*, Gedisa: Barcelona, 1995.

⁵ Cf. P. Martínez Freire «El desafío de las emociones a las ciencias cognitivas», p. 57.

III. DUALISMO EXPLICATIVO, NO DUALISMO ONTOLOGICO

El monismo anómalo intenta también resolver otro problema clásico. Decía más arriba que las descripciones que hacemos de ciertos acontecimientos que son acciones a menudo se expresan en términos causales: los deseos, creencias e intenciones por los cuales describimos las razones de las acciones son sus causas. Pero toda una tradición filosófica, la tradición hermenéutica, así como filósofos como Wittgenstein sostienen lo contrario: las razones no pueden ser causas. Si lo fuesen, afirman tales filósofos, podríamos observar la relación de causalidad que existe entre una intención (un deseo y una creencia) y el movimiento corporal que produce, del mismo modo que observamos la relación causal entre una piedra y el trozo de vidrio roto. Pero las intenciones no pueden ser observadas. Accedemos a ellas en primera persona. Además, si admitimos el principio 2) antes mencionado, allí donde se da una causa, hay una ley estricta. Pero hemos admitido que no hay leyes psicológicas ni leyes psicofísicas estrictas. Por consiguiente, parece —en contra de la hipótesis davidsoniana— que las razones no pueden ser causas. Por tanto, de ahí se puede concluir que nuestras explicaciones psicológicas no son explicaciones causales, son racionalizaciones y explicaciones psicológicas. Habría pues un dualismo fundamental de las explicaciones psicológicas y de las explicaciones físicas.

Nuevamente, se entraría en contradicción con el proyecto de las ciencias cognitivas, al menos con aquél que tiende a un monismo explicativo y pretende mantener una continuidad entre las explicaciones causales naturalistas y las explicaciones psicológicas. Pero si el monismo anómalo de Davidson es correcto, se podría conservar el dualismo explicativo sin adoptar el dualismo ontológico. Nuestras explicaciones mentalistas usuales, efectivamente, son causales, pero no son causales en el sentido de estar apoyadas por leyes estrictas. Y las razones, efectivamente, son causas, cuando las describimos en el vocabulario de la neurofisiología o en el vocabulario cognitivo, si éste es un vocabulario introducido en un nivel que podría reducirse a lo físico. Podemos pues admitir el carácter hermenéutico de las explicaciones establecidas mediante razones, sin por ello suscribir el anti-naturalismo propio de las concepciones hermenéuticas usuales.

Otra razón por la cual el monismo anómalo no estaría comprometido con un dualismo ontológico y no nos llevaría a una división entre dos universos (el de la naturaleza y el de la libertad) es que podemos admitir, según Davidson, la dualidad entre las propiedades mentales y las propiedades físicas sin que ello suponga admitir la inexistencia de relación alguna de dependencia entre las segundas y las primeras. Al contrario, existe tal relación, insiste Davidson, y puede ser caracterizada por medio del concepto de *sobrevenida* de las propiedades mentales con respecto a las propiedades físicas.

Davidson introduce este concepto en su célebre artículo «Mental Events» (1970), para insistir en la idea de que la sobrevenida de lo mental sobre lo físico no es incompatible con la irreductibilidad de las propiedades mentales a las físicas. En términos generales, la noción hace referencia a la relación de dependencia que existe entre una base (en este caso, la de los hechos físicos) y otros hechos (o propiedades) que dependen de la base, pero que tienen ciertas particularidades tales que la reductibilidad de lo que sobreviene a la base no está garantizada. Es decir, las descripciones de acontecimientos psicológicos individuales *sobrevenen* (varían mutuamente y dependen) a las descripciones físicas pero no se reducen a éstas. «Aunque la posición que describo –afirma Davidson– niega que haya leyes psicofísicas, es compatible con la opinión de que las características mentales son en algún sentido dependientes de características físicas o *sobrevenientes* con respecto a ellas. Esta sobrevenida puede ser interpretada como que no puede haber dos acontecimientos que sean iguales bajo todos sus aspectos físicos pero que difieran desde algún aspecto mental. Este tipo de dependencia o de sobrevenida no implica la reductibilidad por medio de leyes o definiciones»⁶.

Dicho de otra manera, si hay diferencias mentales o psicológicas entre dos individuos, hay diferencias físicas entre dichos individuos⁷. Pero esto no implica que todo cambio en una propiedad mental *m* de un acontecimiento mental implicará un cambio en las propiedades *m* que pudiesen poseer otros acontecimientos. Dicho de otro modo, la sobrevenida de lo mental con respecto a lo físico supone la dependencia de lo mental con respecto a lo físico, pero no la reductibilidad de lo mental a lo físico. Esta posibilidad teórica vedaría a un tiempo el dualismo ontológico y el materialismo reduccionista.

IV. IRREDUCTIBILIDAD DE LOS CONCEPTOS MENTALES

¿A qué se debe la irreductibilidad de los conceptos mentales? Ya he esbozado antes una razón: los tipos de leyes que rigen en psicología no son los mismos que los que valen para la física o cualquier otra ciencia reductible a la física. Pero, para Davidson, la raíz de la irreductibilidad procede de otra parte. Pro-

⁶ D. Davidson, «Mental Events», p. 214. Considero más adecuado mantener el término *sobrevenida* para traducir *supervenience*, en lugar de *derivabilidad* que utiliza M. Candel en su traducción al castellano, afín de poder respetar la derivación de los adjetivos *supervenient* (sobreveniente) y *subvenient* (subveniente) del sustantivo, que de lo contrario se perdería.

⁷ Aunque, como señala Jaegwon Kim en *Supervenience and Mind*, el concepto de *sobrevenida* tiene distintas significaciones, éste es el sentido que le da David Lewis, uno de los primeros autores en utilizar esta noción: «hay sobrevenida cuando no hay diferencias de un tipo, si no las hay del otro». Cf. D. Lewis, *On the Plurality of Worlds*. Oxford: Blackwell, 1986, p. 14.

viene de una característica peculiar del vocabulario mental. Las actitudes proposicionales, tales como las creencias, deseos e intenciones no pueden ser poseídas aisladamente: tenemos una creencia sólo si tenemos otras creencias que también ellas se encuentran ligadas a deseos e intenciones. Así yo no puedo atribuirle a alguien la creencia de que esto es una nube, si no puedo atribuirle la creencia de que es un cuerpo opaco, compuesto de vapor de agua, etc.. Se trata de una característica propia de nuestras atribuciones de creencias y de otras actitudes que impone una condición «normativa» sobre estas atribuciones: éstas deben presuponer una multiplicidad, quizá indefinida aunque no infinita, de otros estados del mismo tipo o de tipo diferente y de lazos de coherencia entre los contenidos de estos estados.

En otras palabras, debemos presuponer que las actitudes que nosotros atribuimos tienen una cierta estructura racional (relaciones de coherencia lógica) y una estructura holística (una estructura compleja en red). Por una parte, en la medida en que dicha estructura es racional y normativa, la naturaleza de lo que es atribuido, cuando otorgamos actitudes y conceptos mentales, depende al menos en parte de las propiedades normativas en cuestión, que son las que postula el intérprete o el que las atribuye. Por consiguiente, no es totalmente independiente, en su naturaleza y en su existencia, de las normas de racionalidad de un intérprete. Esto significa, sostiene Davidson, que los objetos mentales –creencias, deseos, etc.– no son entidades independientes de quien las atribuya, sino que existen debido a la interacción entre el intérprete y aquél al que interpreta. Y, por otra, puesto que la estructura en cuestión es holística, no puede ser determinada con toda precisión: si no podemos siempre atribuir una creencia (por ejemplo, que esto es agua) independientemente de otra (por ejemplo, esto es líquido) y de otros estados mentales (por ejemplo, el deseo de beber), esto significa que habrá varios esquemas de interpretación posibles, o si se prefiere, varios esquemas de medida de los conceptos mentales para un sujeto interpretado y, por consiguiente, habrá una indeterminación necesaria de la interpretación. Esta indeterminación no es absoluta, tiene grados; siempre logramos interpretar más o menos las creencias, deseos y otras actitudes de los demás. Pero nunca se dará una interpretación absolutamente correcta, perfecta de lo mental.

Todas estas características –racionalidad, holismo, indeterminación– se alían para hacernos pensar, afirma Davidson, que debe de haber algo específico en los conceptos psicológicos –al menos en aquellos que implican la atribución de actitudes– con respecto a la atribución de propiedades físicas. Estas últimas no responden a las mismas exigencias de racionalidad, coherencia y holismo que las que se imponen para los conceptos psicológicos. Es más, si bien es cierto que podemos hablar con respecto a los conceptos físicos de racionalidad, coherencia u holismo, sin embargo estas características no son de

la misma naturaleza que sus homólogas atribuibles a los conceptos psicológicos. De ahí la irreductibilidad de los dos tipos de conceptos. Cuando medimos las propiedades físicas de los objetos, podemos basarnos en principios que son objetivos y comunes. Pero cuando medimos las propiedades mentales, al interpretar individuos, no podemos apoyarnos en criterios comunes de racionalidad; tenemos que basarnos en nuestros propios criterios de racionalidad y no hay –afirma Davidson– otra objetividad que la que descansa en la intersubjetividad y la comunicación con los demás. Nuestro criterio regulador es el acuerdo. El acuerdo es un ideal en la interpretación de los demás, pero también es un principio de partida inevitable; suponemos que los demás son tan racionales como nosotros, que tienen tantas creencias verdaderas como nosotros, que son tan coherentes lógicamente como nosotros, aunque descubramos que no es así realmente, o que no somos tan racionales o verídicos como pensábamos. Logramos así progresivamente, mediante sucesivos ajustes, alcanzar un acuerdo intersubjetivo en la comunicación. Esto no debe hacernos pensar –apunta Davidson– que los principios que rigen nuestra interpretación de los individuos no sean objetivos, ni que nos impidan compartir un mundo objetivo común. Centrémonos en la interpretación del discurso.

V. INTERPRETACIÓN DE LOS CONCEPTOS MENTALES

Cuando queremos interpretar lo que un individuo dice, ya sea en nuestra lengua o en otra lengua, no disponemos de más datos que las frases que este individuo enuncia, los datos empíricos referidos al entorno exterior y nuestro conocimiento de sus estados mentales. Podría pensarse que uno solo de estos tipos de evidencia empírica bastaría para llegar a los otros dos, pero, en cada caso, esto resulta insuficiente.

No podemos partir únicamente del comportamiento verbal para determinar las creencias, los demás estados mentales y las significaciones de las frases enunciadas, porque –como hemos visto en relación al holismo de lo mental– los datos de la conducta sobredeterminan los contenidos mentales y las significaciones.

Tampoco podemos partir de los datos del entorno, ni de los datos físicos de la cadena causal que va desde los acontecimientos del mundo exterior a los acontecimientos cerebrales del sujeto, porque las cadenas causales sobredeterminan los contenidos mentales y las significaciones. Sin un intérprete que determine desde el exterior cómo una cadena causal que va del mundo a las palabras determina el conocimiento de la significación de una palabra, no hay manera de determinar si el sujeto utiliza esta palabra correctamente o no. Únicamente una comunidad de intérpretes puede determinar el sentido correcto. Únicamente una comunidad de intérpretes puede determinar si un pensamien-

to es en realidad un pensamiento. No hay, en otras palabras, una teoría causal de la referencia de las palabras, ni de la significación, ni de los contenidos mentales.

En fin, no podemos apoyarnos en los datos mentales o psicológicos independientemente de otros tipos de datos, ya que el reconocimiento de un contenido de pensamiento depende del reconocimiento de la significación dada a las palabras. Al menos por lo que se refiere a los pensamientos proposicionales, su contenido no puede ser determinado independientemente del contenido de las frases que los sujetos son susceptibles de enunciar, es decir, de las significaciones lingüísticas⁸. Esto no significa que una creencia sea una disposición a enunciar una frase sino que quiere decir que el contenido de una creencia particular no puede ser interpretado independientemente de una interpretación de las frases del lenguaje de un individuo, como tampoco de sus acciones.

Nuevamente, nos encontramos aquí con el holismo, pero también con la irreductibilidad. Interpretar un sujeto, supone pues tener a un tiempo una teoría de los contenidos de sus pensamientos, una teoría de la significación de sus frases y una teoría de sus acciones, y ninguna puede ser reducida a la otra, ni es primera con respecto a las demás. ¿Cómo debemos entonces proceder en la interpretación, si nos encontramos en una situación de «interpretación radical», en la que ningún contenido interpretado es conocido anticipadamente? Pues, habría que partir –señala Davidson– primero de las frases enunciadas por un individuo. Aquellas que parezcan responder a un acto de asentimiento de un individuo o que él considere como verdaderas, servirán de guía. El intérprete, al confrontar éstas a las frases que él mismo considera verdaderas y al entorno, tratará de maximizar el acuerdo o, más exactamente, minimizar los desacuerdos no explicados. En este sentido, se regirá por un «principio de caridad». Tratará de hacer corresponder el mayor número de frases que él considere verdaderas en su propio lenguaje con el mayor número de frases enunciadas por el individuo que está siendo interpretado⁹. La interpretación del discurso tenderá a producir el mayor número de frases de ese tipo ('S' es verdadero si y sólo si S). Seguidamente, tratará de establecer una estructura en el lenguaje del individuo interpretado. Aquí tampoco hay elección, salvo la de partir de la estructura de su propio lenguaje. Esto supone que él mismo debe tener acceso a la significación de los términos de su propio lenguaje y conocerla. Tiene sobre ella una autoridad en primera persona.

Pero debemos ir más lejos –apunta Davidson– y suponer que la mayor parte de las frases que el individuo interpretado considera como verdaderas lo

⁸ Esto es sin duda una diferencia esencial entre el pensamiento humano y los contenidos mentales que podemos atribuir a criaturas sin lenguaje.

⁹ Así, si el sujeto interpretado enuncia «it's raining» y ocurre que llueve en su entorno, el intérprete podrá emitir la hipótesis de que «it's raining» es verdad si y solo si llueve en el entorno del sujeto.

son efectivamente, al margen de que *nosotros*, intérpretes, las consideremos verdaderas. Por supuesto que cabe el error, pero el error radical, masivo, a juicio de Davidson, no es posible. Esto significa que una condición de posibilidad de toda interpretación es un mundo exterior común y objetivo, una estructura general de verdades; no sólo de verdades para nosotros intérpretes y locutores, sino de verdades independientes de los intérpretes y de los locutores. Esto tiene una consecuencia clara, a saber, que una forma de relativismo, según el cual el mundo es relativo a nuestro esquema conceptual y según el cual podría haber esquemas conceptuales intraducibles los unos a los otros o «incommensurables», estaría injustificada. Según Davidson, en toda interpretación se presupone un mundo común.

En este sentido, nuestro autor suscribe una cierta forma de externalismo, según el cual los contenidos de las frases y de los pensamientos están determinados por un entorno exterior. Este externalismo de los pensamientos y de las significaciones implica entre otras cosas que los pensamientos y sus significaciones no puedan estar determinados por acontecimientos internos a los agentes, fuera de su entorno y de la interpretación que los demás den de sus pensamientos y de sus significaciones. El pensamiento y las significaciones no estarían –a juicio de Davidson– exclusivamente en la cabeza de la gente.

VI. CRÍTICAS A DAVIDSON

Lo que tiene de atractivo el intento de Davidson de compatibilizar el fisicalismo monista y el antireduccionismo, la teoría causal de la acción y el rechazo de las leyes causales estrictas en psicología es precisamente también lo que constituye la raíz de las tensiones inherentes a su posición.

Las críticas que se han hecho al monismo anómalo, en general, se centran en el hecho de que resulta difícil imaginar cómo los acontecimientos mentales pueden ser idénticos a los acontecimientos físicos sin que de ello se siga que existan leyes que unan los dos tipos de acontecimientos. Davidson en efecto no niega que no exista dependencia de los primeros con respecto a los segundos, ni que dos individuos que fuesen totalmente idénticos físicamente deban ser idénticos psicológicamente. Esta dependencia es lo que él denomina la sobrevenida de lo mental sobre lo físico. Pero esta sobrevenida no nos permite explicar la naturaleza exacta de la dependencia de lo mental con respecto a lo físico¹⁰. El problema es que si tal dependencia sin reducción existe, Davidson no nos dice en qué consiste.

¹⁰ B. McLaughlin, «Anomalous Monism and the Irreducibility of the Mental», en E. Lepore y B. McLaughlin, *Action and Events. Perspectives on the Philosophy of Donald Davidson*. Oxford: Blackwell, 1985, p. 365. Considera que la concepción débil de la sobrevenida de Davidson y su monismo anómalo hacen problemático su fisicalismo.

Otra dificultad que se le plantea a la teoría davidsoniana, estrechamente ligada a la anterior, es la de la causalidad mental. Según el monismo anómalo, los acontecimientos mentales causan acontecimientos físicos, pero no los causan en virtud del hecho de que ejemplifiquen leyes psicológicas, puesto que no hay leyes de este tipo, en sentido estricto. Se entiende pues que sólo pueden causarlos en tanto que ejemplifican leyes físicas. Esto significaría que los acontecimientos mentales (por ejemplo creencias y deseos) no causan acontecimientos físicos (por ejemplo movimientos corporales) en virtud de sus propiedades mentales (el hecho de que tengan tal o cual contenido), sino sólo en virtud de las propiedades físicas que ejemplifican. Lo mental sería entonces «inerte» causalmente o epifenoménico¹¹. Davidson no niega en efecto que haya leyes psicológicas y, por tanto, en ese sentido no niega que la noción de causalidad mental no sea independiente de la de ley. Pero éstas leyes psicológicas no son leyes «estrictas»; únicamente son leyes *ceteris paribus*, sometidas a excepciones. Esta dificultad no es, desde luego, exclusiva del monismo anómalo. En la filosofía clásica y contemporánea afecta a otras teorías monistas. Pero la fuerte dualidad que mantiene Davidson entre el nivel ontológico (el de los acontecimientos) y el nivel explicativo (el de las descripciones), así como entre los dos tipos de explicaciones (físicas y psicológicas) pone esta dificultad especialmente de relieve.

Muchas de las objeciones que se le plantean a la teoría davidsoniana de la mente tienen que ver con el holismo de lo mental y han sido especialmente desarrollada por Fodor en *Holism, A shopper's Guide* (1992). Según Fodor, el holismo defendido por Davidson impediría toda concepción realista de la significación y de la mente ya que implica que no podamos establecer las condiciones de individuación del sentido de una frase o del contenido de un pensamiento. Dichas condiciones siempre serán relativas a un conjunto, más o menos definido, y quizá indefinidamente amplio, de otras significaciones o contenidos. Además, señala Fodor, al hacer de la significación y del pensamiento el producto de una interpretación realizada por parte de un observador exterior que atribuye, en tercera persona, contenidos semánticos, reduciría tales contenidos a los contenidos atribuidos. La argumentación de Davidson acerca de la interpretación se aproximaría así a una postura antirealista de lo mental, según

¹¹ Con respecto a las críticas sobre el epifenomenalismo presente en Davidson, véase J. Kim, «Epiphenomenal and Supervenient Causation» y «Postscripts on mental causation» en *Supervenience and Mind*, pp. 92-108 y 358-367. Kim sostiene que el monismo anómalo es en realidad un dualismo tal que no otorga a lo mental ningún trabajo útil y su aparición sigue siendo totalmente misteriosa y causalmente inexplicable (p. 270). Kim reprocha a Davidson que la noción de sobrevenida que utiliza sólo serviría para simular el fisicalismo, hacer el monismo ontológico totalmente indiferente y convertir el dualismo conceptual en fundamental.

la cual se afirmaría que sólo tenemos una mente en la medida en que somos interpretados como tales.

Esta crítica se puede asociar a la anterior: lo mental sólo existiría en tanto que es descrito, pero no en sí mismo; por tanto, es completamente inerte e ineficaz. En este sentido, Madell en *Mind and Materialism* (1986) acusa a Davidson de ignorar la perspectiva subjetiva en primera persona que todos tenemos de los estados mentales. Davidson desarrolla una respuesta a esta crítica en varios de sus trabajos: «First-Person Authority» (1984), «Knowing One's Own Mind» (1987). Además Maddell señala que una cosa es afirmar que sólo podemos comprender a las demás personas si les atribuimos actitudes proposicionales y otra bien distinta afirmar que la raíz de estas actitudes proposicionales reside en nuestra interpretación (p. 100).

Con respecto al holismo y el antirealismo de Davidson, habría que señalar que la posición davidsoniana resultará antirealista, si adoptamos una forma particularmente fuerte de realismo en cuanto a la mente y a la significación. En realidad, la tesis sobre la que se apoyan Fodor y otros filósofos materialistas contemporáneos para sostener que los estados mentales son tan reales como pueden serlo las montañas o que la psicología versa sobre objetos tan sólidos como la geografía o la geología es una intuición que no está asegurada. El monismo anómalo de Davidson y su teoría de la interpretación precisamente se proponen poner de manifiesto la intuición contraria.

REFERENCIAS BIBLIOGRÁFICAS

- DAVIDSON, Donald 1963: «Actions, Reasons and Causes», *Journal of Philosophy*, 60, pp. 685-700, reedit. en *Essays on Actions & Events*.
- _____ 1967: «The Logical Form of Actions Sentences» en N. Rescher, *The Logic of Decision and Action*. Pittsburgh: University of Pittsburgh Press, pp. 81-95. reedit. en *Essays on Actions & Events*.
- _____ 1970: «Mental Events» in L. Foster y J. Swanson, *Experience and Theory*. Amherst: University of Massachussets Press, pp. 79-101, reedit. en *Essays on Actions & Events*. (tr. esp. en *Filosofía de la psicología*).
- _____ 1973a: «The Material Mind» in P. Suppes et al., *Logic, Methodology and Philosophy of Science*, IV. Amsterdam: North Holland, pp. 709-722, reedit. en *Essays on Actions & Events*. (tr. esp. en *Filosofía de la psicología*).
- _____ 1973b: «Psychology as Philosophy» en S. Brown, MacMillan (eds.) in *Philosophy of Psychology*, pp. 41-52, pp. 60-67, reedit. en *Essays on Actions & Events* (tr. esp. en *Filosofía de la psicología*).
- _____ 1980: *Essays on Actions & Events*. Oxford: Oxford University Press, Hay tr. esp. de O. Hansberg, J.-A. Robles y M. Valdés en Barcelona: Crítica, 1995.
- _____ 1987: «Knowing One's Own Mind», *Proceedings and Addresses of the American Philosophical Association*, 60, pp. 441-458 (tr. esp. en *Mente, mundo y acción*).

- ____ 1992: *Mente, mundo y acción*, intr. y tr. de Carlos Moya Espí. Barcelona: Paidós.
- ____ 1994: *Filosofía de la psicología*, intr., tr. y notas de Miguel Candel. Barcelona: Anthropos.
- FODOR, Jerry y Ernest LEPORE 1992: *Holism: A Shopper's Guide*. Oxford: Blackwell.
- GUTTENPLAN, Samuel (ed.) 1994: *A Companion to the Philosophy of Mind*. Oxford: Blackwell.
- HEIL, John y Alfred MELE, (eds.) 1993: *Mental Causation*. Oxford: Oxford University Press.
- KIM, Jaegwon 1993: *Supervenience and Mind. Selected Philosophical Essays*. Cambridge: Cambridge University Press.
- LEPORE, Ernest y B. MCLAUGHLIN, (eds.) 1985: *Action and Events. Perspectives on the Philosophy of Donald Davidson*. Oxford: Blackwell.
- LEWIS, David 1986: *On the Plurality of Worlds*. Oxford: Blackwell.
- MADELL, Geoffrey 1986: *Mind and Materialism*. Edinburgh: Edinburgh University Press.
- MARTINEZ FREIRE, Pascual 1993: *La nueva filosofía de la mente*. Barcelona: Gedisa.
- ____ 2000: «El desafío de las emociones a las ciencias cognitivas», *Thémata*, 25, pp. 55-66.
- MCLAUGHLIN, B. 1985: «Anomalous Monism and the Irreducibility of the Mental», en *Action and Events. Perspectives on the Philosophy of Donald Davidson*.
- VERMAZEN, Bruce y Merrill B. HINTIKKA 1985: *Essays on Davidson. Actions and Events*. Oxford: Oxford University Press.

Alicia Rodríguez Serón es profesora asociada de Filosofía de la mente en la Universidad de Málaga. Autora de *El estudio de la memoria en Bergson: antecedentes e implicaciones filosóficas* (Universidad de Málaga, 1998) y de «Fundamentos metodológicos de la Psicología en Bergson y Freud», *Revista de Historia de la Psicología*, 17 (1996).

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