

Del molino de Leibniz a la brecha de Levine: Sobre mentes, cerebros y máquinas

Rafael Andrés Alemañ Berenguer
Universidad de Alicante

Resumen: La célebre metáfora del molino de Leibniz planteó un problema que mucho después Levine denominó «brecha explicativa», en referencia a la disyunción aparente entre los aspectos mentales y materiales de la realidad. El monismo neutral, a este respecto, se revela como un digno rival del monismo materialista, aceptando la subjetividad irreductible de los *qualia*. Esta tesis separa en cierto modo la psicología de la física y, si a ella unimos una hipótesis adicional, destierra la posibilidad de máquinas pensantes.

Palabras clave: Mente, materia, monismo neutral, materialismo, *qualia*.

From Leibniz's mill to Levine's gap: On minds, brains, and machines

Summary: Leibniz's famous windmill metaphor posed a problem that much later Levine named the «explanatory gap» due to the apparent disjunction between the mental and material aspects of reality. In this regard, neutral monism happens to be a worthy rival to materialist monism by accepting the irreducible subjectivity of *qualia*. This thesis somehow separates psychology from physics and, if we add an additional hypothesis to it, banishes the possibility of thinking machines.

Keywords: Mind, matter, neutral monism, materialism, *qualia*

Recibido: 29 de junio de 2022

Aceptado: 21 de diciembre de 2022

DOI: 10.24310/nyl.18.2024.15073

1. Introducción

El clásico artículo de Levine sobre la dificultad de derivar nuestras más íntimas experiencias subjetivas a partir del conocimiento disponible sobre la base material de los procesos cerebrales, publicado cuatro décadas atrás, marcó un hito en su campo y consagró la expresión «brecha explicativa» (*explanatory gap*). No tanto por haber aportado novedades sustanciales como por la nitidez de sus argumentos, se convirtió desde entonces en texto de referencia para cuantos autores se interesaron por la relación mente-materia. Para los propósitos de este trabajo —y equiparándola al término «conciencia»— adoptaremos como definición de «mente» «[...] la dimensión autotransparente de la vida psíquica, en virtud de la cual el sujeto pensante se convierte en espectador activo de sí mismo, [...]» (Arana, 2015: 19). Por otra parte, para la «materia» convendría «[...] un concepto abierto que designa sin más especificación cualquier entidad que esté inmersa en el espacio-tiempo y acate las leyes naturales [...]» (Ibid: 20).

Los términos del problema apenas han cambiado desde que ese celebrado artículo vio la luz; bien al contrario, sus aristas se revelan hoy más numerosas y afiladas, aunque su núcleo temático permanece intacto. Sigue siendo tan pertinente hoy como entonces abordar las mismas preguntas que Levine dejó abiertas: ¿cuál es el vínculo entre las dimensiones material y mental de la realidad?; ¿son inconmensurables entre ellas?, ¿o ese aparente abismo categorial entre ambas se debe tan solo a nuestros sesgos y prejuicios? La cuestión se enturbió aún más con la llegada de los computadores y la así llamada inteligencia artificial, toda vez que ese nuevo mundo sugirió la posibilidad de que los sistemas electrónicos de suficiente complejidad desarrollaran alguna forma de consciencia parangonable con la mente humana.

El fondo de la cuestión, vestido con diferentes ropajes, siempre gira en torno a la aparentemente insalvable cisura entre la mente y la materia. Los intentos de apuntalar una imagen unificada del mundo por encima de este dilema han sido numerosos y controvertidos, con derivaciones y que llegan

hasta nuestros días. Partiendo en el epígrafe segundo de las consideraciones realizadas desde Leibniz hasta los contemporáneos de Levine, las secciones posteriores confrontarán el monismo materialista y el neutro, cuyas visiones resultan más cercana al conocimiento científico firmemente establecido, con especial hincapié en el monismo neutral —ya en el apartado tercero— menos difundido que su homólogo materialista.

Parte esencial de nuestra discusión será el análisis de la naturaleza de los *qualia* (en los apartados cuarto y quinto), especialísimo e irremplazable componente en la arquitectura de la realidad del que no podemos prescindir si queremos que en ella las mentes jueguen algún papel. Este punto, a su vez, nos llevará a preguntarnos en el epígrafe sexto si cabe atribuir fenómenos mentales a entidades electromecánicas, como los avanzados computadores y autómatas que la tecnología futura nos promete. Y en relación con ello se expondrá una posible formulación de los siempre polémicos vínculos entre ciencias como la física y la psicología, dejando a un lado la apasionante cuestión de la conciencia —y la autoconciencia— cuya extensión desborda los límites de este trabajo. El apartado de conclusiones, en séptimo lugar, cerrará la exposición con un breve resumen de las tesis más relevantes aquí defendidas.

2. De Leibniz a Levine

Pocos dudarían que la piedra de toque en la construcción de la propia identidad viene fundamentalmente dada por la distinción entre la intimidad de nuestro yo y la existencia de un mundo externo omnipresente, incontestable y siempre ajeno a los deseos humanos. Nuestros pensamientos, anhelos y temores —nuestra mente, al cabo— quedan al margen de la fría impavidez del mundo material que nos envuelve, tan abrumador como ineludible. Con la triste excepción de los desórdenes mentales graves, cualquier persona es consciente de la taxativa disyunción entre nuestra vida mental y los acontecimientos materiales del entorno. Desde tan privilegiada atalaya los seres humanos nos entregamos a explorar hasta el últi-

mo rincón del universo, hacia los reinos de lo obnubilantemente grande y de lo inimaginablemente pequeño. Las ciencias de la naturaleza, y aun las ciencias sociales, dieron buena prueba de nuestro éxito en esas lides. Pero quedaba todavía un receptáculo por expungar, cual era nuestra propia mis-midad. Allí, en las mismas raíces de nuestro universo interior acaso residiese la respuesta a la incógnita de la esencia, la existencia y —por qué no— la trascendencia del ser humano.

Pero la empresa resultó inesperadamente ardua, superando con mucho la pugna intelectual que condujo al descubrimiento de las leyes de Newton o la evolución darwiniana. Parecía haber algo radicalmente distinto en el mundo mental que lo apartaba sin remedio de los aspectos materiales de la realidad. No bastó con abandonar las apelaciones a un alma inmaterial y concentrar las investigaciones en el cerebro, requisito indispensable para descubrir hasta donde cabía llegar por ese camino en la explicación naturalista de nuestros cimientos psicológicos. La frontera entre mente y materia persistía, invisible e impenetrable, resistiendo todo esfuerzo de traspasarla. El gran Leibniz fue uno de los autores que más temprana y diáfana-mente señaló el núcleo del problema con su celebrada metáfora del molino:

Hay que reconocer, por otra parte, que la *percepción* y lo que de ella depende resultan *inexplicables per razones mecánicas*, esto es, por medio de las figuras y de los movimientos. Porque, imaginémosnos que haya una máquina cuya estructura la haga pensar, sentir y tener percepción; se la podrá concebir agrandada, conservando las mismas proporciones, de tal manera que podamos entrar en ella como en un molino. Esto supuesto, una vez dentro, no hallaremos sino piezas que se impelen unas a otras, pero nunca nada con que explicar una percepción. Así, pues, esto hay que buscarlo en la sustancia simple, no en lo compuesto o en la máquina. Más aún, no cabe hallar en la sustancia simple otra cosa excepto esto, es decir, excepto las percepciones y sus cambios. Y solamente en esto también pueden consistir todas las *Acciones internas* de las sustancias simples. (Leibniz, 1981, p. 87)

Lo que se ha conocido también como la «brecha de Leibniz» (*Leibniz's gap*) situó en el centro del escenario la cuestión medular del dilema mente-materia. Con este argumento, el filósofo alemán trataba de mostrar las limitaciones del mecanicismo materialista a la hora de explicar la naturaleza de nuestra vida mental, la cual deviene ininteligible en el marco de esta doctrina según el argumento precedente. Por un lado, desde una perspectiva meramente material solo tenemos acceso a los aspectos objetivos de la realidad (propiedades medibles, cambios de estado), a través de los cuales intentamos desvelar los rasgos estructurales básicos del mundo físico. En el extremo opuesto, aparentemente, se halla nuestra subjetividad interna, en virtud de la cual no solo componemos las teorías abstractas que representan el universo material, sino que asimismo nos vemos atravesados por recuerdos, pensamientos y emociones de toda clase.

Es de gran importancia subrayar que la dicotomía señalada por Leibniz supone un reto tanto para dualistas como para monistas. Quienes defiendan que mente y cerebro son dos entidades separadas habrán de explicar el modo en que interactúan, a menos que se abandonen a una suerte de armonía preestablecida o paralelismo psicofísico —una escapatória demasiado artificiosa para ser tomada en serio. Y cuantos aboguen por el monismo deberán justificar la escandalosa disparidad entre el plano material y el mental que damos por descontado en todas nuestras descripciones del mundo.

Con la añadidura de alguna sofisticación formal, argumentos análogos fueron esgrimidos muchos años después por Saul Kripke, Thomas Nagel, Frank Jackson, Joseph Levine y David J. Chalmers en sus célebres estudios al respecto. Kripke (1971, 1979, 1980) desarrolló una versión mucho más sofisticada del argumento cartesiano favorable a una distinción bien fundada entre la mente y el cerebro. Apoyándose en la semántica de mundos posibles y en su teoría de la referencia directa para términos lingüísticos, este filósofo concluyó que los estados mentales no tenían necesariamente que identificarse con estados físicos de esa delicada porción de materia que llamamos cerebro. El eje argumental sostenía que, por ejemplo, el concepto físico de calor como agitación molecular era distinto de nuestra sensación

de calor, mientras que no cabe diferenciar entre «sentir dolor» y «la sensación de sentir dolor». De nuevo, como advertía Leibniz, la insondable cisura entre objetividad y subjetividad.

En efecto, parece haber un conflicto irreconciliable entre, por una parte, los colores, sonidos y olores que captamos –sin olvidar el flujo interno de nuestros pensamientos– y por otra el inane mundo de partículas y campos descrito por la física fundamental que sustenta todo ello (Sellars, 1962). Esa fue la tensión esencial que Levine recogió en su ya clásico artículo consagrando la expresión *explanatory gap*. Con esa locución, tan breve como contundente, logró condensar los escollos en que embarrancaba la metafísica materialista cuando pretendía dar cuenta de los aspectos cualitativos de los estados mentales:

Obviously, there is something right about it. Indeed, we do feel that the causal role of pain is crucial to our concept of it, and that discovering the physical mechanism by which this causal role is affected explains an important facet of what there is to be explained about pain. However, there is more to our concept of pain than its causal role, there is its qualitative character, how it feels; and what is left unexplained by the discovery of C-fiber firing is why pain should feel the way it does. For there seems to be nothing about C-fiber firing which makes it naturally «fit» the phenomenal properties of pain, any more than it would fit some other set of phenomenal properties. Unlike its functional role, the identification of the qualitative side of pain with C-fiber firing (or some property of C-fiber firing) leaves the connection between it and what we identify it with completely mysterious. One might say, it makes the way pain feels into merely a brute fact. (Levine, 1983: 357)

Contemplar la faceta mental de algunos procesos neuronales como meramente un hecho bruto (*brute fact*) de la naturaleza, sin necesidad de explicaciones ulteriores, nos dejaría en una situación un tanto desairada. Levine concede que los fenómenos mentales parecen surgir tan solo cuando

el sistema nervioso alcanza cierto grado de complejidad, lo que resulta sin duda verosímil. No obstante, de ser así nos encontraríamos con la dificultad de que precisamente en niveles superiores de complejidad no esperaríamos encontrar esa clase de hechos en bruto. Más aún, Levine prosigue señalando que, aun cuando conociésemos la correlación entre ciertas configuraciones físicas del cerebro humano y las experiencias internas vividas por el sujeto poseedor, no sabríamos justificar por qué tales configuraciones implican tales experiencias y no otras:

The point I am trying to make was captured by Locke in his discussion of the relation between primary and secondary qualities. He states that the simple ideas which we experience in response to impingements from the external world bear no intelligible relation to the corpuscular processes underlying impingement and response. Rather, the two sets of phenomena – corpuscular processes and simple ideas – are stuck together in an arbitrary manner. The simple ideas go with their respective corpuscular configurations because God chose to so attach them. He could have chosen to do it differently. Now, so long as the two states of affairs seem arbitrarily stuck together in this way, imagination will pry them apart. Thus, it is the non-intelligibility of the connection between the feeling of pain and its physical correlate that underlies the apparent contingency of that connection. Another way to support my contention that psycho-physical (or psycho-functional) identity statements leave an explanatory gap will also serve to establish the corollary I mentioned at the beginning of this paper; namely, that even if some psycho-physical identity statements are true, we can't determine exactly which ones are true. The two claims, that there is an explanatory gap and that such identities are, in a sense, unknowable, are interdependent and mutually supporting. (Ibid: 359)

Esta brecha explicativa –una sima más bien– nos coloca en la incómoda tesitura de reconocer que seríamos incapaces de determinar si una criatura con una estructura física distinta de la nuestra podría verse sometida a las

mismas experiencias en caso de recibir los mismos estímulos. De hecho, ni siquiera tendría sentido comparar nuestra sensibilidad con la suya (si la tuviese):

Now, if there were some intrinsic connections discernible between having one's C-fibers firing (or being in functional state F) and what it's like to be in pain, by which I mean that experiencing the latter was intelligible in terms of the properties of the former, then we could derive our measure of similarity from the nature of the explanation. Whatever properties of the firing of C-fibers (or being in state F) that explained the feel of pain would determine the properties a kind of physical (or functional) state had to have in order to count as feeling like our pain. But without this explanatory gap filled in, facts about the kind or the existence of phenomenal experiences of pain in creatures physically (or functionally) different from us become impossible to determine. (Ibid: 360)

3. La respuesta del materialismo monista

Son pocos los autores que en los últimos tiempos han osado preconizar cualquier forma de dualismo (Eccles, 1992) o de pansiquismo (Torday y Miller, 2018), si bien algunos físicos cuánticos profesan –al menos en sus momentos más desmadrados– una versión del idealismo que dejaría en mantillas al obispo Berkeley. Lo cierto es que los triunfos de la neurociencia a lo largo del siglo XX han sido tan apabullantes que la indiscutible correlación entre el sistema nervioso central y nuestra vida psíquica ha dejado expedito el camino para la preponderancia del monismo materialista. La réplica de los adeptos al materialismo monista ha consistido tradicionalmente en rechazar la existencia de esa brecha explicativa y, por tanto, negar el problema desde su misma raíz (Churchland, 1996; Papineau, 2002; Dennett, 2005).

En su opinión, los fenómenos mentales se identifican con procesos materiales, en virtud de lo cual nos es dado afirmar que la sensación de dolor es idéntica a la activación de las fibras neuronales de tipo C, o que la visión del

color es otro modo de nombrar cierta actividad en el córtex visual (Place, 1970; Smart, 1990). Así se expone la identidad psiconeural en su versión materialista, enriquecida posteriormente con la más amplia visión del «físicismo no reduccionista», que abraza la posibilidad de que los fenómenos mentales sean resultados emergentes, surgidos a partir de propiedades físicas sin identificarse con ninguna de ellas en particular (Bunge, 1998). Comoquiera que sea, en algún sentido los fenómenos mentales deben identificarse con algún evento material pues de otro modo no podríamos explicar las cadenas causales que parecen existir entre ambos, como sucede en los desórdenes psicosomáticos.

Los seguidores del monismo materialista destacan que –según la ciencia nos ha mostrado– la mente es el nombre otorgado a una colección de funciones cerebrales, de cambios de estado o procesos en los sistemas encefálicos de organismos con cierta complejidad mínima. Esos procesos se hallan estrechamente vinculados al comportamiento observado de tales organismos, sin que haya sustancias mentales de tipo alguno por las que debamos preocuparnos. Como lo que tenemos en realidad es un organismo que opera de una determinada forma, al separar la función del organismo cometemos un grave error categorial. El estómago y los intestinos –por ejemplo– forman el sistema digestivo porque en ellos tiene lugar una serie de procesos bioquímicos que llamamos digestión. Pero más allá del funcionamiento del sistema digestivo, la digestión no es una entidad en sí misma. Por ello nadie suele argüir que en la naturaleza haya digestiones sin sistemas digestivos que las lleven a cabo.

Así pues, el monismo materialista considera que el origen de la aparente brecha explicativa señalada por Levine se encuentra en la resistencia a aceptar esa identidad entre mente y cerebro. Dicho brevemente, sería nuestra adhesión inconsciente al dualismo psicofísico la que nos inhabilitaría para percatarnos de que la brecha de Levine no existe en realidad. El monismo materialista no ignora la pertinencia de preguntarse por qué ciertos procesos físicos llevan aparejados aspectos mentales, o por qué tales

procesos engendran determinadas sensaciones conscientes. Lo que sostienen los materialistas es que esos interrogantes no dependen de una inexistente disparidad ontológica entre estados mentales y estados materiales, algo que cualquiera podría advertir si no estuviésemos cegados muy a menudo por un compromiso implícito con el pensamiento dualista.

¿Cuál sería el origen de tan persistente adhesión al dualismo? Ciertas conjeturas en psicología fisiológica sugieren la posibilidad de una doble vía en nuestra arquitectura cognitiva diseñada para distinguir entre procesos materiales y mentales (Bloom, 2004). Por un lado tendríamos el módulo «mentalista» —por decirlo así— que nos invita a atribuir estados mentales a otros agentes aparentemente intencionales, y por otro actuaría el módulo «materialista» permitiéndonos afrontar los desafíos del mundo físico. En el requisito práctico de que ambas redes cognitivas nunca se activen a la vez —prosigue Bloom— podría hallarse la razón de que concebamos la mente y la materia como dos reinos inconmensurables. Sin embargo, no precisamos suscribir enteramente esta tesis, pues la posible existencia de esos dos cauces no descarta por principio que mente y materia en verdad presenten aspectos radicalmente distintivos, y que nuestro sistema cognitivo se limite a reproducir ese hecho en su estructura.

De un modo u otro, en este debate siempre aparece una arista que el monismo materialista se empeña en ignorar en tanto carece de explicación para ella, posiblemente porque tal explicación resulta inasequible. El filo que la explicación materialista trata de esquivar disimulando sus contorsiones con más o menos fortuna reside en el apercibimiento inmediato de la mente sobre sus propios contenidos. Esa dimensión autorreflexiva que constituye la más pura subjetividad y el centro de nuestra vida psíquica encierra la clave que perseguimos: «[...]. La autotransparencia produce un tipo de recursividad muy peculiar que se da de un modo ubicuo en la mente humana, y de un modo claramente restringido en cualquier otra criatura de la que haya noticia. [...]» (Arana, op. cit: 39).

Aquí se revela el punto crucial del dilema mente-cerebro, cuya importancia es de tal calibre que apenas hay lugar para la exageración por mucho que se insista en él. Ocurre que la subjetividad humana conserva para sí un reducto que se insinúa inexpugnable ante cualquier tentativa de conquista mediante el único instrumento disponible para la racionalidad científica, esto es, el escrutinio objetivo de aquello que investigamos. Ahora bien, ¿cómo abordar objetivamente algo que por su propia índole es irrenunciablemente subjetivo? Dicho brevemente, «[...] objetivar la mente es transformar en objeto lo que de por sí es sujeto, y por tanto negarlo en cuanto tal. [...]» (Arana, op. cit: 126).

Si la mente humana se halla ligada al cerebro —y todo indica que así es— se colige que habrá surgido gradualmente por el mismo proceso de evolución biológica que dio lugar a la progresiva encefalización de nuestra especie. Y de ello se deduciría que también cabe atribuir cierto tipo de estados mentales a otros vertebrados con un sistema nervioso relativamente complejo (Proctor et al, 2015; Feinberg y Mallatt, 2016a, 2016b; Birch, 2019; Ginsburg y Jablonka, 2019). No poseerán las capacidades de recursividad, abstracción y autoconciencia típicas de los humanos, pero sin duda tampoco resulta descabellado hablar de mente en ciertos animales, aun cuando se separe de la nuestra por una barrera cualitativamente infranqueable.

La pertinencia de interrogarse por los estados mentales de algunos animales conduce a plantarse si también tales estados o procesos poseerán las cualidades intrínsecas que únicamente nos es dado conocer a nosotros por medio de la introspección, pues corresponden al aspecto puramente subjetivo de nuestra intimidad psíquica. Esta duda inspiró el célebre artículo de Thomas Nagel (1974) en el que se preguntaba sobre la percepción del mundo de un mamífero tan distinto de nosotros como un murciélago. Así nos vemos conducidos al centro de una polémica tan añeja como profunda acerca de esas experiencias internas que conforman el tejido de nuestra vida mental, los *qualia*.

4. El problema de los *qualia*

El término *qualia* –plural de *quale*– se debe al estadounidense Clarence Irving Lewis, quien lo introdujo en 1929 para referirse al carácter fenoménico de las experiencias provocadas por los datos sensoriales (Crane, 2000), esa aprehensión directa, intrínseca y no representacional de nuestros estados mentales que se expresa como «ser tal que» (*what is like*, en inglés). Algunos autores se remiten a ese carácter inefable y privado para rechazar su existencia (Dennett, 1993), mientras que otros se acogen al argumento sobre lenguajes privados de Wittgenstein (1952) para declararlos imposibles (Scruton, 2016). Pero la refutación de los lenguajes privados ofrecida por Wittgenstein solo prueba que la conciencia directa de nuestros estados mentales no basta para determinar el significado de los términos utilizados para referirnos a ellos. Así pues, algo tan patente como el hecho de la autotransparencia que caracteriza nuestro mundo mental interno solo puede combatirse mediante esforzadas piruetas argumentativas como las desplegadas por Scruton:

[...]. Atribuimos cualidades a muchas de nuestras experiencias no mirando adentro, sino mirando afuera, a las cualidades secundarias de los objetos. Ver el rojo es una experiencia visual clara; pero describir esa experiencia es describir qué aspecto tienen las cosas rojas, lo que a su vez requiere mostrarlas. Las cosas rojas son cosas como esta; y ver el rojo es una experiencia visual que tienes cuando ves algo como esto. Ver el rojo es distinto de ver el verde, porque las cosas rojas son distintas de las cosas verdes. Sin duda, eso plantea la cuestión de las cualidades secundarias: ¿están de verdad ahí, en las cosas que parecen poseerlas? Me inclino a pensar que las cualidades secundarias son disposiciones a suscitar experiencias en el observador normal, pero que las experiencias se deben identificar a su vez por medio de las cosas que percibimos. [...]. (Ibid: 54)

Nótese que aquí el filósofo británico deposita toda la fuerza de su argumento en la presunta ausencia de *qualia* referidos a experiencias puramente

internas, sin correlato exterior, suposición contra la cual milita una constelación de evidencias en nada desdeñables. Cuando un individuo queda en un estado de aislamiento sensorial, la suspensión de casi cualquier dato de los sentidos procedente del exterior se suple con una actividad espontánea del sistema nervioso central en forma de alucinaciones. Aun cuando esas alucinaciones carezcan de conexión con la realidad externa al sujeto, no cabe dudar que ese mismo sujeto las experimenta y forman parte de sus estados mentales en esos momentos con el acceso directo a ellas que tal situación implica. Luego la sugerencia de Scruton, según la cual la mente depende enteramente de su enlace con el mundo circundante, no se sostiene.

Nueva munición —y de grueso calibre— se aportó a la batalla argumental cuando la imaginaria Mary salió de su no menos imaginaria habitación. En este célebre experimento mental, propuesto por Frank Jackson (1982, 1986), una hipotética Mary ha nacido y crecido en el interior de una estancia donde se dan todas las tonalidades del gris entre el blanco y el negro, aprendiendo todo cuanto cabe saber sobre los aspectos físicos de la luz y la cromaticidad, pero sin percepción directa de color alguno. Cuando Mary salga al exterior y vea por sí misma los colores parece indudable que experimentará un estado mental antes desconocido para ella. De ser así, se deduce que los estados mentales poseen componentes —los *qualia*— que no se pueden explicar en términos únicamente físicos (como el tipo de conocimiento que Mary tenía sobre los colores antes de abandonar la estancia). La réplica de Dennett (1991), arguyendo que un conocimiento completo de la teoría física permitiría deducir incluso propiedades como los *qualia*, suena tan poco creíble que no se sabe si su propio autor la abrazaba sin reservas.

Pocos negarían la radical diferencia existente entre conocer la luz como una propiedad física caracterizada por la magnitud que llamamos «longitud de onda», y la percepción de la luz en cualquiera de sus posibles colores. Naturalmente, eso no significa que la mente sea una entidad por completo independiente de la materia, al estilo dualista; más bien revela la esencial distinción cualitativa entre el conocimiento abstracto y el directo. El saber abstracto nos lo proporciona el reino de la física teórica con sus propieda-

des primarias (masas, cargas, longitudes de onda, etc.), mientras que en el conocimiento directo participa un aspecto fenoménico, perceptual, que entraña un ingrediente –los *qualia*– del que carece la otra modalidad de conocimiento. Por eso se diría que nos lastra una insuperable dificultad de principio en el intento de identificar sucesos mentales con eventos materiales (Fodor, 1975).

Es cierto que la ciencia porfía por explicar la realidad más allá de las apariencias en el sentido fenoménico, aunque el acaecimiento de tales «apariencias» sea una cuestión científicamente interesante en sí misma. De hecho, la ciencia moderna germinó cuando sus cultivadores concedieron primacía a las propiedades primarias –no directamente perceptibles– relegando las propiedades secundarias al casillero fenoménico, es decir, a un estrato puramente subjetivo de la realidad. Pero ahí nos precipitamos de nuevo en el garlito del que no acabamos de salir: la investigación científica, como exploración objetiva del mundo, no emplea predicados fenomenistas (salvo, quizás, como derivados), lo que se torna una pretensión imposible de satisfacer cuando abordamos la subjetividad misma del individuo, inmediata a la experiencia e irreducible a términos externalistas.

El carácter estrictamente privado e inefable de los *qualia* parece situarlos allende cualquier verificación empírica, a diferencia del alto grado de precisión y fiabilidad de los mecanismos neurofisiológicos. Sin embargo, a mediados del siglo XX el austriaco Herbert Feigl sugirió que la descripción psicológica y la neurológica no diferían demasiado, puesto que tanto la designación de la experiencia privada como la designación de los objetos del mundo físico consisten en traducir una sensación a un cierto lenguaje (Feigl, 1967). Este planteamiento de la «doble designación» (en contraste con la conjetura de Bloom, propia del monismo materialista, sobre una duplicidad de módulos cognitivos) nos deja ya a las puertas de un monismo no materialista, cuya aseveración primordial defiende que la mente y la materia son dos aspectos dispares de una misma realidad subyacente, prescindiendo de cualquier apelación a instancias más allá del mundo natural. Estamos hablando del monismo neutral.

Formulado por William James a finales del siglo XIX y promovido activamente por Bertrand Russell en el XX, el monismo neutral ha sido siempre cuestionado por la dificultad de concretar cuál es el elemento básico de la realidad cuya diversa estructuración, según nuestra perspectiva, da lugar a las categorías de «mente» y «materia». James presentó las sensaciones como ese ingrediente constructivo, elección recogida inicialmente por Russell hasta que el filósofo británico se decantó por las cadenas de eventos espacio-temporales. En tal caso estaríamos justificados al preguntarnos cómo los eventos espacio-temporales —un concepto rotundamente físico— pueden constituir una mente. ¿No sería esto una modalidad disfrazada de fisicismo y, a la postre, de materialismo?

5. La propuesta del monismo neutral

Para absolvernos de la acusación según la cual el monismo neutral tan solo encubre una versión vergonzante del monismo materialista, basta con empezar aceptando algo obvio para muchos: en los *qualia* nos encontramos con elementos irreductiblemente subjetivos de la realidad (Searle, 1992), que surgen como propiedades emergentes en sistemas tan complejos como el cerebro humano y nos permiten construir marcos categoriales definitorios de la materia y de la mente (Alemañ, 2019). Esa interioridad y auto-transparencia de los *qualia* establece un límite en la capacidad explicativa del materialismo fisicalista. A finales del siglo XX se popularizó el término «misterismo» —*mysterianism*, en inglés— para asignárselo a quienes sospechaban que la mente humana nunca acabaría de despejar del todo las brumas que envuelven la consciencia de sí misma que tanto la distingue del resto del universo (McGinn, 1989; Flanagan, 1991). En otras palabras, se supone que nunca resolveremos el llamado «problema duro de la consciencia» (Chalmers, 1995), que consiste en la búsqueda de algún puente para sortear la brecha de Levine. Pero este apelativo comporta una cierta injusticia, por cuanto podría hacerse extensivo a cualquier campo del conocimiento. Nunca obtendremos respuestas definitivas cualquiera que sea el

problema científico que abordemos, con la diferencia de que en este caso la existencia de fronteras se manifiesta de forma más imperiosa. Pese a ello, nada nos impide presionar sobre tales fronteras para conseguir que retrocedan lo máximo posible; cualquier otro comportamiento sería una traición al espíritu científico que debe animar estas indagaciones.

El primer punto a tener muy en cuenta en cualquier aproximación al monismo neutral es que no se trata en modo alguno de un mero disfraz, más o menos hábil, para el dualismo tradicional. Estamos tan acostumbrados a pensar en términos dicotómicos, entre el dualismo y el monismo materialista, que no solemos reparar en la posibilidad de alguna otra forma de monismo no materialista. Y sin embargo, autores que no suscriben abiertamente esta tercera opción se expresan en ocasiones mediante un muy atinado lenguaje neutralista:

Un matiz que debe tenerse en cuenta para bieninterpretar mi posición es que *no sostengo que cuerpo y conciencia remitan a realidades disjuntas*, sino a una misma realidad abstraída a partir de dos criterios diferentes (sometimiento a la legalidad natural y autotransparencia), criterios que no se dejan reducir uno a otro, aunque tampoco supone el primero la mera negación del segundo, ni agotan necesariamente entre ambos las posibilidades de la realidad abstraída. [...]. Hay aspectos de la conciencia [...] que muy bien pueden ser corpóreos (esto es, sometidos a la legalidad natural) aunque la constitución de la relación sujeto-objeto (esto es, el surgimiento de un espacio interior de representación autotransparente) en modo alguno es explicable por la ciencia natural ni cualquier otro instrumento cognitivo, pues al fin y al cabo todo estos resultan de ella. (Arana, op. cit: 174-175; cursivas en el original)

Es decir, cuando agrupamos las secuencias de eventos espacio-temporales en obediencia a las leyes físicas tenemos un proceso que cabe llamar «material». Pero cuando entran en juego aspectos como los *qualia* o la autotransparencia de nuestra interioridad psíquica —resistentes a una descripción puramente física— nos las vemos con procesos mentales.

[...] la conciencia y el resto de lo que constituye la realidad del hombre (cuerpo, organismo o lo que sea) no se relacionan entre sí como dos sustancias diferentes, sino como aspectos o dimensiones de una realidad única. Si tomamos esa realidad tal cual y la hacemos pasar por el filtro naturalista, obtendremos una versión parcial de lo humano que se atienen estrictamente a esquemas legaliformes. Podemos llamar «*cuerpo*» a su contenido [...]. Por otra parte, «*conciencia*» no quiere decir [...] sino, de modo paralelo, aquellos aspectos de la realidad humana que conseguimos atrapar cuando prescindimos de todo lo que no sea el mundo interior de representaciones autotransparentes. [...]. (Ibid, 192; cursivas en el original)

Así, el significado de la proposición «ver un objeto de color verde» puede considerarse el resultado de una secuencia causal que comienza en una superficie capaz de reflejar luz de una cierta longitud de onda. El haz de luz llega a nuestros ojos y activan las células especializadas de su interior que transmiten la correspondiente señal eléctrica a lo largo del nervio óptico hasta llegar a la región cerebral donde será descodificada mediante la excitación de los oportunos grupos neuronales. Ahora bien, junto con la perspectiva meramente física –objetiva– del suceso coexiste un aspecto mental por completo subjetivo y por ello estrictamente privado. Esa vertiente mental, a diferencia de la descripción física común para todos, depende en exclusiva de cada individuo. Y por si ello fuera poco, los *qualia* se revelan aquí como un componente imprescindible de la experiencia mental que denominamos «ver un objeto de color verde».

En un sentido material, «ver un objeto de color verde» se simbolizaría como el triplete $\langle C, S, O \rangle$, donde C es la secuencia de eventos espacio-temporales causalmente relacionados que conectan la superficie externa con la zona encefálica de la que depende la visión del color; S es el sistema material –el cerebro– en que termina la secuencia C , y O representa cualquier observador capaz de realizar una descripción físicamente objetiva del proceso. La dimensión mental de ese mismo hecho vendría simbolizada por

el cuarteto $\langle C, S, O^*, Q \rangle$ en el que las diferencias las marcan la presencia de los *qualia*, Q , y la prescripción de que solo un individuo concreto O^* tiene acceso inmediato a su propio estado mental cuando ve el color verde en cuestión.

6. Física y psicología. La cuestión de la máquina pensante

Preguntarnos por el vínculo entre la mente y la materia comporta cuestionarnos, antes o después, la conexión entre las leyes de la física y las leyes —si existen como tales— de la psicología. Constatamos las regularidades objetivas que manifiestan entidades consideradas elementales en el mundo físico. Las pautas que rigen los cambios en las partículas aparentemente básicas (electrones, fotones, quarks, neutrinos) son buenas candidatas a ese rango, sin olvidar que esas entidades fundamentales se desenvuelven en un espacio-tiempo cuyo estatuto ontológico sigue siendo objeto de disputa.

Todo lo existente existe de alguna forma concreta; no se puede simplemente «ser» sin más. Esos modos de existencia son las propiedades de las cosas, las cuales nos permiten explicar, en cierto sentido, las interacciones y los cambios de las entidades fundamentales. Las propiedades fundamentales se limitan unas a otras estableciendo compromisos más o menos permanentes —por ejemplo, los principios de conservación— que justifican la estabilidad del mundo real. Las ecuaciones que describen las diversas interacciones entre partículas representan otro ejemplo de ello, en este caso expresando un patrón dinámico de relaciones entre esos objetos fundamentales. Así pues, una ley fundamental de la naturaleza sería una relación permanente entre propiedades básicas de entidades elementales.

Ahora bien, en la naturaleza no solo tenemos partículas elementales dispersas; hay rocas, virus, animales y plantas, planetas y estrellas, galaxias y todo tipo de objetos materiales compuestos por esas entidades básicas. Los entes básicos de la realidad poseen la capacidad de asociarse para formar entidades compuestas más complejas, sistemas, que representamos mediante modelos abstractos y conceptos tan generales como el de sistema. Nunca

sobra insistir en que un sistema no es una mera colección de elementos; más allá de eso, los sistemas poseen una estructura y una composición características que los dotan de propiedades —o «formas de ser»— diferentes de las exhibidas por sus componentes individuales. Tales son las propiedades emergentes que caracterizan los distintos niveles de complejidad de los sistemas reales.

Parece lógico suponer que estas propiedades emergentes, a su vez, obedecerán sus propias leyes, las cuales, debido a su propia especificidad no podrán ser completamente deducidas a partir de las leyes de los niveles inferiores. Piénsese en el caso de la viscosidad de un fluido; no cabe hablar con sentido de la viscosidad de las moléculas individuales que constituyen el fluido, pero tampoco puede dudarse que el fluido como tal posee esa propiedad. La viscosidad es una propiedad emergente que depende en parte del tipo molecular, aunque no se reduzca enteramente a él. No obstante, el punto clave aquí consiste en advertir que en el ámbito físico los sistemas materiales dan lugar a diversas propiedades emergentes, según el nivel de complejidad en que nos movamos, cuyo carácter también es estrictamente físico.

Caso distinto parece ser el de la psicología. La actividad del cerebro como sistema material ciertamente posee propiedades emergentes (como las ondas alfa) de las que carecen las neuronas individuales, pero ocurre que surgen asimismo otras, como los *qualia*, que no pueden interpretarse en términos fisicistas. Esta es la peculiaridad —o, al menos, una de las peculiaridades— que nos permite distinguir la categoría mental de la material, aun cuando ambas remitan a una misma realidad subyacente. Es verdad que de este modo la psicología ocuparía un puesto especial dentro de las ciencias, al exhibir algunas propiedades emergentes sin parangón con el resto, pero este rasgo tan solo debería preocupar al materialista y no al monista neutral consecuente.

Una faceta más vino a añadirse al problema cuando el desarrollo de la microelectrónica, aplicada a la tecnología de la computación durante la segunda mitad del siglo XX, alentó el sueño de una entidad consciente —con

inteligencia y sentimientos análogos a los humanos— formada por cables y transistores en lugar de neuronas y glía. El vertiginoso aumento de la potencia de los computadores reforzó las esperanzas de culminar el llamado «programa fuerte de la Inteligencia Artificial» aunque nadie aclaraba cómo podía obrarse el prodigio de que un mosaico de circuitos comenzase a pensar y sentir. John Searle (1980), con su celebrada e imbatida metáfora de la «habitación china» puso el dedo en la llaga de la distinción entre operaciones sintácticas y contenidos semánticos; es decir, las computadoras pueden manipular símbolos, pero no atribuir significados como sí hace la mente humana. La réplica de los Churchland (1990) careció de consistencia por cuanto en última instancia confundía las propiedades emergentes de carácter físico y las de carácter psicológico.

Con toda seguridad, la controversia en este ámbito perdurará durante largo tiempo aún, lo que no debería impedirnos, a la luz de la evidencia hoy disponible, formular una hipótesis ontológica susceptible de acomodarse sin excesivas tensiones en el seno del monismo neutral. Los *qualia* parecen constituir un componente irrenunciable de la consciencia humana, aunque no el único (Peters, 2014), dado que la capacidad semántica (la semiosis, en el sentido de Pierce, Morris y Saussure) se presentan asimismo como otro constituyente indispensable. Ciertamente que los *qualia* son elementos cognitivos diferentes de los contenidos semánticos, pues estos últimos poseen un carácter conceptual que no tienen los primeros; sin embargo, dado que ambos configuran dos de las dimensiones esenciales de la mente humana parece irresistible la tentación de atribuir su origen al carácter singular de dicho ámbito.

Es decir, tanto la autotransparencia —manifestada en los *qualia*— como la capacidad semántica de la vida psíquica aflorarían, ambas y exclusivamente, en biosistemas de suficiente complejidad (sin entrar a discutir ahora qué sería una complejidad «suficiente») para desarrollar lo que denominamos mente humana. A la que bien podríamos llamar «hipótesis psicógena» debemos la conjetura según la cual, si aceptamos la aseveración de Searle según la cual un dispositivo electrónico —con independencia de su sofisti-

cación— carece de mente puesto que solo los biosistemas complejos disfrutan de capacidad semántica, concluiremos que tampoco poseerá qualia en la medida en que los consideremos asimismo emergentes específicos de la actividad neural.

La hipótesis psicógena propuesta sostendría, pues, la existencia de un lazo indisoluble —con muchos de sus detalles aún por esclarecer— entre autotransparencia y semiosis en la mente humana (y presumiblemente en otras no humanas de comparable complejidad). Esto se superpone a la afirmación de que solo los biosistemas conformados por moléculas orgánicas que alcancen suficiente grado de complejidad pueden dar origen a propiedades emergentes específicamente mentales (como los *qualia*) y a contenidos semánticos. Y de ello, en consecuencia, se inferiría que albergan mentes.

Sin duda, esta hipótesis sobre la génesis de la mente —no solo la humana— concede un especial privilegio al carbono, cuyas combinaciones moleculares construirán los biosistemas en los que se desarrollará el psiquismo. Sin embargo, este punto no necesariamente debe interpretarse como un desdoro, sino más bien como una llamada de atención sobre nuestra todavía profunda ignorancia acerca de las potencialidades aun insospechadas de la materia, no solo de la mente.

7. Conclusiones

La rabiosa rotundidad con que se presenta ante el hombre el hiato entre la dimensión mental y la corporal de su existencia ha suscitado uno de los problemas filosóficos más añejos e inexpugnables en la historia del pensamiento humano. El dualismo mente-materia, tan natural para los primeros pensadores como para el individuo común, quedó postergado por la revolución científica que dio paso a la Modernidad. Impulsado por esos vientos de cambio, el monismo materialista se enseñoreó del campo de batalla, despojando a sus rivales siquiera de la licitud de participar en la liza. Y los espectaculares progresos de las neurociencias en el siglo XX aparentaban

respaldarlo, pues la identidad psiconeural, que igualaba la actividad nerviosa con la vida mental del sujeto, parecía ahuyentar cualquier otra opción que no fuese la puramente materialista.

Con todo y ello, declinar la adhesión al monismo materialista no implica abandonarse en los brazos de un dualismo que de principio se sitúa extramuros de la indagación científica. Este dilema se ve superado por la existencia del monismo neutral, que concibe mente y materia como dos expresiones de un mismo estrato profundo de la realidad que se manifiesta ante nosotros en esas dos modalidades. La identidad psiconeural de los materialistas se preserva, aunque sin conceder a la materia una primacía ontológica no exenta de reveses.

Los autores que desde Leibniz a Levine, y posteriores, han señalado la disparidad de categorías entre la conciencia y la corporeidad plantearon un reto, aún vigente, que el monismo materialista no solo no ha respondido sino que a menudo se ha negado a reconocer. El monismo neutral, aun cuando tampoco posee una respuesta definitiva para un problema que acaso carezca de ella, ofrece una nueva perspectiva y propuestas del mayor interés. En los anteriores apartados se han esbozado algunas de ellas, comenzando por la admisión de los *qualia* como propiedades emergentes irreductiblemente subjetivas —y, por tanto, elementos básicos e irrenunciables de la consciencia— a los que se une en la mente humana la capacidad de crear y adjudicar contenidos semánticos.

Tales consideraciones nos condujeron a las dos hipótesis principales defendidas en este trabajo, a saber: (1º) El privilegio de los biosistemas complejos basados en el carbono como los únicos en los que se puede manifestar la doble dimensión mente-materia defendida por el monismo neutral, lo que excluye la posibilidad de máquinas con una mente genuina. (2º) En ese ámbito mental superior la conciencia humana se constituye a partir de elementos como los *qualia* y la capacidad semántica, entre otros posibles, cuyas conexiones siguen siendo temas abiertos a la investigación.

A partir de todo ello comprenderemos la futilidad de empeños como el «programa fuerte de la inteligencia artificial». La psicología adquiere así un

estatuto intelectual radicalmente distinto al del resto de las ciencias, toda vez que en ella arraiga al menos un componente –los *qualia*– (y quizás también la semiosis) no del todo reducible a términos, emergentes o no, meramente físicos. Pero esta inaccesibilidad no debe llevarnos al desaliento; más bien supone un nuevo acicate para espolearnos una vez más en la búsqueda de los verdaderos límites del conocimiento humano.

Bibliografía

Alemañ, R. (2019): «El monismo neutral y la superación de la dualidad mente-materia», *Naturaleza y Libertad. Revista de Estudios Interdisciplinarios* (11), pp. 11–33

Arana, J. (2015): *La Conciencia Inexplicada. Ensayo sobre los límites de la comprensión naturalista de la mente*. Madrid: Biblioteca Nueva.

Birch, J. (2019): «In Search of the Origins of Consciousness», *Acta Biotheoretica* (68), pp. 287–294.

Bloom, P. (2004): *Descartes' baby*. New York: Basic books.

Bunge, M. (1998): *El problema mente-cerebro. Un enfoque psicobiológico*. Madrid: Tecnos.

Chalmers, D. J. (1995): «Facing up to the problem of consciousness», *Journal of Consciousness Studies*, (2), pp. 200–219.

Churchland, P. (1996): «The rediscovery of light», *Journal of Philosophy* (93), pp. 211–28.

Churchland, P; Churchland, P. (1990): «Could a machine think?», *Scientific American* (262), nº 1, pp. 32–39

Crane, T. (2000): «The origins of qualia», en T. Crane y S. Patterson (eds.), *History of the Mind-Body Problem*. Routledge: London, pp. 169–194.

Dennett, D. (1991). *Consciousness Explained*. Boston: Little, Brown and Co

Dennett, D. (1993): «Quining Qualia», en A. I. Goldman (ed.), *Readings in Philosophy and Cognitive Science*. Cambridge (Mass): MIT Press, pp. 381-414.

Dennett, D. (2005): *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge: MIT Press

Eccles, J. (1992): *La evolución del cerebro: creación de la conciencia*. Barcelona: Labor.

Flanagan, O. (1991): *The Science of the Mind*. Cambridge (Mass): MIT Press.

Feigl, H. (1967): *The Mental and the Physical: The Essay and a Postscript*. Minneapolis: University of Minnesota Press.

- Feinberg, T; Mallatt J. (2016a): *The Ancient Origins of Consciousness: How the Brain Created Experience*. Cambridge (Mass): MIT Press.
- Feinberg, T; Mallatt J. (2016b): «The Nature of Primary Consciousness. A New Synthesis», *Consciousness and Cognition* (43), pp. 113–127
- Fodor, J. (1975): *The Language of Thought*. Cambridge (Mass): Harvard University Press.
- Ginsburg, S; Jablonka, E. (2019): *The Evolution of the Sensitive Soul: Learning and the Origin of Consciousness*. Cambridge (Mass): MIT Press.
- Jackson, F. (1982): «Epiphenomenal Qualia», *Philosophical Quarterly* (32), nº 127, pp.127–136.
- Jackson, F. (1986): «What Mary Didn't Know», *Journal of Philosophy* (83), nº 5, pp. 291–295.
- Kripke, S. (1971): «Identity and necessity», en M. Munitz (ed.), *Identity and Individuation*. New York: New York University Press, pp. 135–164.
- Kripke, S. (1979): «A puzzle about belief», en A. Margalit (ed.), *Meaning and Use*. Dordrecht: Reidel, pp. 239–283
- Kripke, S. (1980): *Naming and Necessity*. Cambridge (Mass): Harvard University Press.
- Leibniz, G. W. (1981): *Monadología*. (Edición trilingüe. Introducción de Gustavo Bueno. Traducción de Julián Velarde). Oviedo: Pentalfa
- Levine, J. (1983): «Materialism and qualia: the explanatory gap», *Pacific Philosophical Quarterly* (64), pp. 354–361.
- McGinn, C. (1989): «Can we solve the mind-body problem?», *Mind* (98), pp. 349–366.
- Nagel, T. (1974): «What is it like to be a bat?», *Philosophical Review* (83), pp. 435–450.
- Papineau, D. (2002): *Thinking about Consciousness*. Oxford: Oxford University Press.
- Peters, F. (2014): «Consciousness Should Not Be Confused With Qualia», *Logos and Episteme* (5). 10.5840/logos-episteme20145123.
- Place, U. (1970): «Is consciousness a brain process?», en C. Borst (ed.), *The Mind-Brain Identity Theory*. London: Palgrave, London, pp. 42–51.
- Proctor, H; Carder, G; Cornish, A. (2013): «Searching for Animal Sentience: A Systematic Review of the Scientific Literature», *Animals: An Open Access Journal* (3), nº. 3, pp. 882–906.
- Romero, G. (2018): *Scientific Philosophy*. New York: Springer.
- Scruton, R. (2016): *El alma del mundo*. Madrid: Rialp

Searle, J. (1980): «Minds, Brains and Programs», *Behavioral and Brain Sciences* (3), nº 3, pp. 417–457,

Searle, J. (1992): *The Rediscovery of Mind*. Cambridge (Mass): MIT Press

Sellars, W. (1962): «Philosophy and the Scientific Image of Man», en R. Colodny (ed.), *Frontiers of Science and Philosophy*. Pittsburgh: University of Pittsburgh Press, pp. 35–78.

Smart, J. (1999): «Sensations and Brain Processes», en N. Warburton (ed.), *Philosophy: Basic Readings*. London: Routledge, pp. 281 – 294

Torday, J; Miller, W. (2018): «The Cosmologic Continuum from Physics to Consciousness», *Progress in Biophysics & Molecular Biology* (140), pp. 41–48.

Wittgenstein, L. (1952): *Philosophical Investigations*. Oxford (UK): Oxford University Press, pt. 1, sec. 220 y ss.

Rafael Andrés Alemañ Berenguer
raalbe.autor@gmail.com

