



La existencia de los qualia: Una presentación crítica.

The existence of qualia: A critical presentation.

DOI: 10.32870/sincronia.axxvii.n83.2b23

Alejandro Villamor Iglesias

IES Rosalía de Castro (ESPAÑA)

CE: alejandrovillamoriglesias@yahoo.es / ID ORCID: 0000-0001-7265-7528

Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial 4.0 Internacional](https://creativecommons.org/licenses/by-nc/4.0/)

Recibido: 08/01/2023

Revisado: 08/03/2023

Aprobado: 04/04/2023

RESUMEN

El presente artículo tiene como objetivo principal proporcionar al lector una breve aproximación a los *qualia*. La utilidad de esta empresa es aportar un texto básico en español sobre algunas polémicas que giraron en torno a dicho concepto. Para ello, el trabajo analiza críticamente la cuestión de su existencia a través de dos debates paralelos. El primero corresponde con el intento de Dan Dennett de demostrar la incoherencia del concepto de *quale* tal y como es presentado por los dualistas. En segundo lugar, se introduce el debate entre Frank Jackson y Paul Churchland al respecto del experimento mental de la habitación de Mary del primero.

Palabras clave: Filosofía de la mente, qualia, Daniel Dennett, argumento de la habitación de Mary.

ABSTRACT

The main purpose of this article is to provide the reader with a brief approach to *qualia*. The usefulness of this undertaking is to provide a basic text in Spanish on some of the polemics that surround this concept. To this end, the paper critically analyzes the question of its existence through two parallel debates. The first corresponds to Dan Dennett's attempt to demonstrate the incoherence of the concept of *quale* as presented by dualists. Secondly, the debate between Frank Jackson and Paul Churchland regarding the former's Mary's room thought experiment is introduced.

Keywords: Philosophy of mind, qualia, Daniel Dennett, Mary's room argument.



El proyecto eliminativista de los *qualia* de Dennett

La intuición sobre el cambio de juicios

Enmarcado en un proyecto eliminativista, Dan Dennett se propuso desmontar el mito acerca de la existencia de los *qualia*. Los *qualia* (en plural), en términos de Asier Arias, son “cada una de las cualidades subjetivas diferenciables en un estado mental dado” (Arias, 2021, p. 63). El sabor del café, la sensación del miedo a las alturas o la sensación cromática del azul son todos ellos un *quale* (en singular). Al menos en su sentido tradicional, los *qualia* son un duro obstáculo en el camino hacia una teoría de la mente naturalizada. Estas propiedades son consideradas por muchos autores (cfr. Levine, 1983 o Chalmers, 1999) como óbice para el materialismo o cualquiera perspectiva del mundo en tercera persona (Dennett, 1995) pues esta dejaría algo queda por explicar.

Dennett mantiene, así, en su conocido artículo “Quinear los *qualia*” (Dennett, 2003) la inexistencia de tal obstáculo¹ para dar vía libre al proyecto consistente en la supeditación de la filosofía de la mente a las ciencias empíricas. Como punto de partida, resulta de especial relevancia atender a la caracterización de los *qualia* por parte de Dennett. ¿De qué se pretende negar la existencia? Lo que no se quiere negar bajo ningún concepto, y se pone bastante énfasis en ello, es la existencia de unas propiedades de la experiencia consciente que son experimentadas de una determinada manera. Esto no se niega en ningún momento. Más bien, Dennett pone en cuestión la existencia de una suerte de propiedades especiales de segundo orden con respecto a esa experiencia fenoménica, los *qualia*, que poseerían una serie de características: la infabilidad, la intrinsecabilidad, la privacidad y la accesibilidad o aprehensibilidad directa o inmediata en la conciencia.

La tesis de Dennett no es sólo que las diversas nociones de los *qualia* sean vagas y equívocas, sino que la misma intuición o concepción preteórica de las mismas está errada. “*Qualia*” no refiere a nada, y, para mostrarlo, en este trabajo la finalidad es la extirpación de dicha intuición en sus

¹ La de Dennett se erige, así como una postura que podemos denominar como fuerte en lo que a la crítica de los *qualia* refiere. El error de otros críticos con los *qualia*, afirma, ha sido consistido en asumir de algún modo su existencia. Se han dedicado, estos, al ataque de todos los argumentos pergeñados por los defensores de dichos *qualia* cuando lo que se debería hacer es extirpar la raíz misma del problema. Para Dennett, no existe ningún *quale* (Dennett, 2003, p. 220).



mismos términos, esto es, a través de otras intuiciones. Para ello, ha generado un interés especial el “*generador de intuición 7*”, relativo al célebre caso de los catadores de café Chase y Sanborn. Estos señores, tras estar varios años trabajando para la misma empresa, comentan mutuamente que el sabor de sus cafés ha dejado de gustarles. Aspecto de gran relevancia es que para el resto de la plantilla de la empresa el sabor del café sigue siendo el mismo, algo que acordarán Chase y Sanborn. El problema es dilucidar si lo que ha variado, como propone Chase, son sus gustos o “actitudes reactivas” hacia determinados *qualia* o, como dice Sanborn, el cambio se encuentra en el propio sistema perceptivo. Lo primero que se encarga de señalar Dennett es que, por las dudas manifestadas por nuestros dos sujetos, la introspección llevada a cabo se muestra ineficaz. Los *qualia*, por ende, ya comienzan a no parecer tan accesibles al sujeto epistémico. De las tres posibles soluciones posibles para cada caso —básicamente que, o han variado los *qualia* de los individuos, o sus gustos, o bien sucede algo intermedio— ninguno de los sujetos puede optar con seguridad por ninguna.

Alguien podría considerar la búsqueda de la confirmación de una de las posibilidades a través de pruebas empíricas. Una de ellas podría poner a prueba, por ejemplo, la capacidad de Chase para la identificación de sabores en cataduras a ciegas. U, por ejemplo, la realización de mediciones fisiológicas. Empero, la prueba empírica, en última instancia, tampoco permite alcanzar una solución al problema. Esto es ilustrado por Dennett en el “*generador de intuición 8: la gradual recuperación postoperatoria*”: imaginemos que una intervención quirúrgica modifica las conexiones de las papilas gustativas de Chase de modo que, tras la operación, por ejemplo, el azúcar le sabe salado. Con el tiempo, sin embargo, Chase se adapta a la inversión de un modo tal que su conducta es indiscernible de la de cualquiera otra persona. El problema que nos topamos es análogo al del espectro invertido (cfr. Byrne, 2020): no es posible determinar si la adaptación se produce porque sus *qualia* se han restablecido tal como eran antes de la operación o si, por lo motivos que fueren y al margen de la creencia del sujeto, sus *qualia* no se han “normalizado”. En consecuencia, ni por introspección ni por medios empíricos, según Dennett, podemos alcanzar solución definitiva. Aunque señala:



Lo sorprendente de esto no es sólo que los métodos empíricos no lograrían distinguir lo que parecen ser afirmaciones muy diferentes sobre los *qualia*, sino que no lograrían hacerlo *a pesar de constituir mejor evidencia que las convicciones introspectivas de los propios sujetos* (Dennett, 2003, p. 236).

Las consecuencias que se siguen de ello pondrían en un serio aprieto a esa noción tradicional de *qualia* presentada anteriormente. Estas supuestas propiedades fenoménicas que son los *qualia* ni siquiera, a pesar de la intuición común, serían inmediata o directamente accesibles por los propios sujetos. El motivo es que, como pretende ilustrar el ejemplo de Chase y Sanborn, no somos verdaderamente conscientes de la constancia o cambios que en ellos se producen. Por esto, continúa Dennett, ni el juicio de Chase ni el de Sanborn deberían ser considerados superiores o más importantes que los de cualquier examen en “tercera persona” (Dennett, 2003, pp. 237-238).

Asimismo, Dennett hace especial énfasis en los problemas que tiene la supuesta intrinsecidad de los *qualia*. Este cuestionamiento queda patente con el precedente ejemplo, pero es especialmente claro en el “*generador de intuiciones 9: el bebedor de cerveza experimentado*”. En resumidas cuentas, en este caso busca mostrar que las supuestas propiedades especiales defendidas por los qualófilos son, en realidad, “extrínsecas y relacionales”. La variación entre el primer sorbo de cerveza, no comúnmente disfrutado, y un posterior trago, que provocaría un cierto deleite, mostraría el papel vital que ocupan, por ejemplo, la “actitudes o reacciones” que uno tiene ante sus propias experiencias. El gusto adquirido hacia la cerveza lo muestra. Hasta aquí tenemos, pues, un caso, el de los catadores de café, que comprometería seriamente la intrinsecidad, la accesibilidad inmediata —puesto que los individuos no pueden decidir qué ha sucedido—, la infabilidad —de la que dice que es cuestión de práctica el alcanzar una mayor capacidad para expresar nuestras experiencias (Arias, 2011, p. 21)— y la privacidad —las pruebas empíricas tienen un mayor peso que la propia opinión—.

La propuesta de Dennett podría ser reprochable en al menos algunos aspectos. En primer lugar, como indica Asier Arias, es cuestionable la embestida de Dennett a la intrinsecidad. El



estadounidense parte del supuesto de que la concepción tradicional incorpora esta característica, mas esto se puede poner en duda: “Cabe pues dudar del supuesto —implícito en la argumentación de Dennett, y decisivo para el éxito de la misma— según el cual la intrinsecalidad forma parte tanto de la concepción tradicional como la cotidiana o intuitiva” (Arias, 2011, p. 18). Como muestra Dennett, esta es una característica que compromete especialmente a los *qualia*. Pero su atribución es tensar demasiado la cuerda. No se pone en duda el carácter relacional de esas propiedades subjetivas que serían los *qualia*. Parece que los ejemplos relativos al sabor de la cerveza o el “*generador de intuiciones 10: el experimento mundial de eugenesia*” dan buena cuenta de las sumas dificultades con que se toparán los defensores de ese carácter intrínseco, esto es, aquellos que defienden que “los *qualia* son propiedades que los estados mentales tienen en sí mismos” (Arias, 2011, 4), independientemente de todo lo demás (incluidos los propios juicios de los sujetos). Resulta de enorme dificultad negar que el sabor de la cerveza desde el primer sorbo hasta el último (suponiendo que haya habido varios) haya variado en virtud de determinadas circunstancias relacionales o extrínsecas. Incluso, en la actualidad, parece que existen evidencias empíricas que avalan lo equívoco de esta característica. Por ejemplo:

Los miembros de tribus no familiarizadas con la representación bidimensional de estructuras tridimensionales no perciben dichas representaciones del mismo modo que los sujetos acostumbrados a semejante tipo de representación gráfica, es decir, carecen de los estados mentales (conscientes y adquiridos) que conformarían el contexto que propiciaría el señalado modo de percepción (Arias, 2011, pp. 5-6).

Se pone en duda que los que los *qualia* posean esta propiedad. Así, si nos limitamos a aceptar la posesión de esta característica por parte de los *qualia* con Dennett, entonces estaremos de acuerdo con él en que resulta insostenible. Ahora, otro cantar resulta si convenimos que la noción tradicional de *qualia* puede prescindir de la susodicha característica. Observemos, entonces, el argumento de Dennett tomándonos la licencia de prescindir de esta comprometida característica.



Un primer punto a nombrar se encuentra en la presuposición por la cual Dennett desecha la posibilidad de la infalibilidad de los *qualia*. Tras haber expuesto las diferentes posibles respuestas al problema de Chase y Sanborn, afirma:

No conozco a nadie que hoy en día se sienta tentado a defender la postura extrema de la infalibilidad o la incorregibilidad que establecería que la opción (a) es —y debe ser— la verdad en ambos casos, porque las personas simplemente no pueden equivocarse con respecto a estas cuestiones privadas y subjetivas (Dennett, 2003, p. 230).

Quizás al decir esto se esté dando demasiado por supuesto. Dirá después, sin embargo, que lo que se debe hacer es “tratar a los *qualia* como construcciones lógicas a partir de los juicios de los sujetos sobre sus *qualia*” (Dennett, 2003, p. 231). Nosotros somos, al fin y al cabo, los que, juzgándolos, damos origen a los propios *qualia* al modo el que un novelista determina el color del pelo de uno de sus personajes. La supuesta infalibilidad de los *qualia* se muestra, según esta idea, como una ilusión arbitraria. Pero, ¿es esta una razón de peso que nos obligue a descartar tan pronto la infalibilidad de los *qualia*? ¿No podría ser que Chase estuviera en lo correcto al decir que variaron sus gustos puesto que, aunque no lo recuerda, en el pasado observó como una rata muerta se encontraba en un café de Maxwell House? ¿No puede ser que Sanborn esté en lo correcto cuando afirma que sus *qualia* se han alterado puesto que, aunque no lo relacione, el año pasado sufrió una quemadura en su lengua que afectó a sus papilas gustativas?

Lo que intentamos indicar con esto no es que, por pura casualidad, en estos dos casos los sujetos estén en lo cierto. Más bien, queremos recalcar que Dennett no aporta tampoco ningún motivo por el cual estos dos señores no deban estar en lo cierto. ¿Por qué deberíamos aceptar que nuestros propios juicios son los “actos constitutivos” de lo que llamamos *qualia*? Dennett parece asumir sin más que la infalibilidad de los *qualia* es falsa. Si por infalibilidad entendemos esa característica en virtud de la cual nadie puede corregir la experiencia que el sujeto experimenta, entendemos que ni los propios sujetos, ulteriormente, pueden corregir sus propias experiencias. Uno experimenta lo que experimenta sin más, como estaría conforme Dennett, que no pone en



cuestión la existencia de las sensaciones fenoménicas. Lo que afirma Dennett es que el sujeto en primera persona está huérfano de criterio, más allá de su falible (porque la memoria lo es) convicción.

Imaginemos el siguiente caso: Joaquín es uno de los mejores enólogos del mundo. Todo aquel que lo conoce está de acuerdo en que Joaquín posee un léxico privilegiado para describir todo aquello que saborea. Supongamos que un buen día Joaquín decide probar un determinado vino y, mientras lo saborea, escribe un informe en el cual describe prolijamente sus sensaciones gustativas. Al año siguiente, en un concurso, Joaquín vuelve a probar el mismo vino (él lo sabe). Sin recordar absolutamente nada del informe del año anterior, nuestro enólogo vuelve a escribir un informe del mismo vino. Él cree que el vino le sabe igual, pero, para comprobarlo, decide comparar los dos informes descritos. Resulta que las dos descripciones son prácticamente (excepto en rasgos insignificantes) idénticas.

Dando la vuelta a lo dicho por Dennett, en este caso suponemos que son los juicios los que emanan tras los *qualia*. El mentado sería un caso, no nomológica ni metafísicamente imposible, en el cual el individuo en primera persona podría apelar a un criterio propio, no en tercera persona, para corroborar su criterio. Con esto se muestra un caso en el que el criterio del individuo resulta más sólido, pero la esencia sigue siendo la misma que los casos que Dennett cuestiona. Ahora, lo que deberíamos preguntarnos es ¿debemos aceptar que un grupo de neurocientíficos emplearan las pruebas que emplearan, tendrían más potestad que el criterio del propio Joaquín?

Otro pilar sobre el que se asienta la noción de *qualia* que maneja Dennett es la accesibilidad. Del ejemplo de los catadores de café deriva Dennett la conclusión de que el supuesto acceso directo y privilegiado a los propios *qualia* se puede poner en cuestión por posibles alteraciones en la memoria. Sin embargo, pudiera ser que aquí el de Boston estuviera confundiendo la accesibilidad directa e inmediata con la comparación en la memoria (Arias, 2011, pp. 19-21). Una cosa es nuestra capacidad (o incapacidad) para comparar experiencias de un pasado lejano, de uno más corto o del presente, y otra que yo pueda acceder directamente —no por ningún tipo de inferencia ni modo indirecto— a aquello que ahora estoy experimentando. Una cosa es que Chase no sepa si el café le



sigue sabiendo igual ahora que hace 6 años, y otra es que, al tomar ahora el café, Chase no esté aprehendiendo inmediatamente el *quale*.

Por inefabilidad entendemos la característica gracias a la cual los supuestos *qualia* no serían comunicables en un sentido pleno. De la caída de la característica de la intrinsecalidad se sigue, como las piezas de dominó, la de la inefabilidad. Si de la intrinsecalidad se deriva la atomicidad e inanalizabilidad y, asimismo, estas últimas son las causantes de la inefabilidad, entonces sin intrinsecalidad no tendría por qué haber inefabilidad. Ahora bien, dice Dennett: “¿Qué decir entonces de la inefabilidad? ¿Por qué parece que nuestras experiencias conscientes poseen propiedades inefables? Porque, en efecto, tienen propiedades inefables *en la práctica*” (Dennett, 2003, p. 250). Para Dennett, la inefabilidad no es una barrera que imponga la existencia de los *qualia* como unas propiedades de segundo orden absolutamente privadas e intransferibles, sino, en cualquier caso, una incapacidad práctica. Así, con respecto a esta práctica, tomada al menos ella misma, no alcanzamos a ver con claridad que Dennett aporte ningún ataque demoledor. Él mismo dirá:

Quizá no soy capaz de describir la propiedad o de identificarme en relación con algún rasgo público al que pueda recurrir fácilmente (todavía), pero estoy familiarizado con ella en un grado modesto: puedo referir a la propiedad que detecté: es la propiedad que detecté en ese suceso (Dennett, 2003, p. 252).

Parece que lo que en realidad hace Dennett es dejar la puerta abierta a una superación de la cuestión de la inefabilidad. Pero nada allende esta posibilidad.

Nos resta dedicar algunas palabras a la última de las características que sostienen, para Dennett, el concepto tradicional de *qualia*. Esta es la privacidad. Como señalamos anteriormente, para Dennett no hay nada que justifique una mayor autoridad de los individuos en perspectiva de primera persona que cualquier prueba en tercera. Esta privacidad no es más que una ilusión de los propios sujetos, dado que ni ellos mismos podrían, como es el caso de Chase y Sanborn, tener claros sus propios *qualia*. Ahora bien, insistiendo en lo ya dicho, a partir del artículo de Dennett parece



dudosa la capacidad de los individuos de discernir la continuidad o variaciones de sus *qualia*, pero no que dejen de ser suyos. Como señala Nagel en su celeberrimo “What Is it Like to Be a Bat?” (Nagel, 1974), parece que por mucha información que tengamos acerca de la ecolocalización no podemos saber cómo es (la sensación de) tenerla.

En conclusión, Dennett se propone demoler la concepción tradicional de *qualia*, de la cual nos da él mismo su caracterización. Si aceptamos esta última, incluida la intrinsecalidad, parece que el argumento de Dennett es irreprochable. Ahora bien, siguiendo a Asier Arias, la puesta en entredicho de su noción tradicional de *qualia* no termina de derrumbar la noción misma de los *qualia* —que seguiría resultando problemática para una teoría naturalizada de la mente—. De lo que se trataría de defender, desde una posición materialista, por tanto, no es la eliminación de esas nociones que nos resultan tan intuitivas como son los *qualia*, sino más bien su incorporación a las teorías naturalistas. El que en la actualidad no se contemple todavía el modo en que esto se pudiera llevar a cabo no significa que, con el progresivo desarrollo de las ciencias, no sea posible.

La “invariabilidad funcional”

Hemos visto en el precedente apartado como Dennett apela al “generador de intuiciones” del catador de vinos para mostrar el carácter relacional de esos supuestos estados mentales que son los *qualia*. En el mismo trabajo, Dennett apela a la “invariabilidad funcional” con un objetivo similar. Es decir, tanto la homogeneidad como la atomicidad son, según Dennett, supuestas consecuencias que se derivan de la intrinsecalidad de los *qualia*. Así, el filósofo estaría aquí intentando socavar la afirmación de que hay algo así como unas propiedades homogéneas y atómicas a lo largo del tiempo. Estas supuestas “propiedades simples o atómicas” son susceptibles de variación, por lo que tal atomicidad, en realidad, se desvanece. El ejemplo propuesto por Dennett para iluminar esto consiste en el adiestramiento del oído para la música (Dennett, 2003, pp. 256-257). Un determinado sonido que en un principio puede semejar atómico, indivisible, con un poco de afinamiento del oído (diseccionando lentamente “el” sonido inicial) terminaría por descubrirse como compuesto. Con este ejemplo busca presentar una de las maneras en las que lo que aparentemente es intrínseco o,



si se quiere, atómico para un sujeto, el tiempo puede mostrar “compuesto y descriptible”. El sonido que en un primer momento parece atómico puede terminar desvelándose como compuesto para el mismo sujeto. Esto es hacia lo que apunta Dennett cuando habla de la “invariabilidad funcional” y de su plasticidad a través del tiempo.

El objetivo de Dennett nuevamente es atacar el concepto de *qualia*. Los frentes abiertos con la invariabilidad son varios. La intrinsecidad, la inmediata accesibilidad o la inefabilidad son algunas de las características de los *qualia* que aquí se tornarían problemáticas. La supuesta “homogeneidad” o “uniformidad” de los *qualia* no respondería más que a incapacidades de los sujetos. En el caso anterior, la incapacidad del lego en música para distinguir un sonido compuesto de uno atómico. Aunque el sonido, desde luego, sea el mismo. Dennett sostiene tras mostrar el ejemplo: “La homogeneidad y la inefabilidad de la primera experiencia han desaparecido, y han sido reemplazadas por una dualidad tan “directamente aprehensible” y claramente descriptible como la de cualquier cuerda” (Dennett, 2003, p. 257). Nuevamente, la estrategia es reducir los *qualia* a disposiciones reactivas, extrínsecas.

La pelota ahora se encuentra en el tejado del qualófilo. Este pudiera responder a Dennett negando que la intrinsecidad sea una propiedad atribuible a los *qualia*. De este modo, el problema se solucionaría pues no negaría la posibilidad de que los *qualia* variaran o se configuraran relacionalmente. En este caso, lo que podría decir este amigo de los *qualia* es que el primer sonido y el último no producen los mismos *qualia*. Puede admitir las variaciones extrínsecas o relacionales sin por ello prescindir de la accesibilidad inmediata, inefabilidad o privacidad del *quale*. En el caso del sonido compuesto también se sostendrían estas características pues, en el momento de percibirlo, serían dos sonidos en lugar de uno. Se mantendría tanto la accesibilidad inmediata a los *qualia* producidos por la guitarra, como la inefabilidad y privacidad.

Asunto un tanto distinto es aquel al que se enfrentara el defensor de la intrinsecidad de los *qualia*. De algún modo, este debería argüir que el *quale* que surge de la guitarra es el mismo, tanto el del primer sonido como el del último. Para esto se podría resguardar con la afirmación de que, sea como fuere, el sonido sigue siendo el mismo y no tendríamos por qué pensar que el *quale* no lo



sea. Esta perspectiva, no obstante, se volvería contra el amigo de los *qualia* puesto que, mientras haya una persona que afirme que la sensación ha variado, la supuesta infalibilidad desmoronaría lo dicho. Aunque no debemos perder de vista, desde luego, que, como apunta el mismo Dennett, hay varias maneras en las que los defensores de los *qualia* caracterizan la intrinsecalidad. El qualófilo podría afirmar que lo que sucede con la educación del oído, en casos como este, es que el sujeto adquiere un modo más complejo de acceder al *mismo quale*. Y aunque alguien afirmase que la sensación del primer sonido al segundo ha variado, el amigo de los *qualia* que siguiese esta senda como respuesta podría decir que se ha accedido al mismo *quale*, en ese caso, de un modo más complejo.

El argumento epistémico de Jackson

En aras de presentar un ataque contra el fisicalismo, proponiendo como posición alternativa un dualismo de propiedades, Jackson presentó el famoso experimento mental de María. Junto con otro experimento muy semejante, el de Fred, el hombre que puede ver colores vetados al resto de nosotros, el de María es presentado en un artículo, “Epiphenomenal Qualia” (1982). En “Epiphenomenal Qualia” Jackson presenta toda una serie de argumentos que, afirma, ponen en serio compromiso la postura fisicalista. Estos son: el argumento del conocimiento, el argumento modal —explotado por Chalmers con su experimento de los zombis (cfr. Chalmers, 1999)— y el argumento “como es ser” (Jackson, 1982). Concretamente, al poner en entredicho la tesis epistémica de que todo el conocimiento es reducible al conocimiento de los hechos físicos, el experimento de María es un argumento epistémico. Jackson intenta mostrar por medio de este que el fisicalismo es falso. Para ello, el experimento se pretende erigir como un “argumento válido a partir de premisas altamente plausibles —aunque, hay que admitirlo, no demostrables—, premisas para la conclusión de que el fisicalismo es falso” (Jackson, 1986, p. 295).

El experimento de María consiste sucintamente en lo siguiente: María es una brillante científica que ha desarrollado su formación académica y su vida en una habitación en blanco y negro. Todo, incluida ella misma, es en blanco y negro. Suponiendo que esto fuera posible, María



sabe absolutamente todo lo que físicamente se puede saber acerca de las personas y de su entorno. La física, la química y la neurofisiología son disciplinas que María conoce en su totalidad. Resulta que un buen día, María sale de su habitación y se encuentra con un mundo lleno de colores que nunca había experimentado. María ha conseguido atesorar, así, un nuevo conocimiento del que carecía. De un modo más formal:

- (1). El fisicalismo sostiene que el mundo es exclusivamente físico.
- (2). Todo conocimiento es conocimiento del mundo.
- (3). El fisicalismo sostiene que todo conocimiento es conocimiento físico.
- (4). El fisicalismo sostiene que el conocimiento físico completo es conocimiento completo del mundo.
- (5). Mientras estaba en su habitación, María tenía un conocimiento físico completo acerca de las personas.
- (6). Percibir cómo es algo nuevo es aprender algo nuevo acerca de las personas.
- (7). Al salir de la habitación, María percibe algo nuevo.
- (8). Al salir de la habitación, María aprende algo nuevo acerca de las personas.
- (9). María no sabía todo lo que se puede saber acerca de las personas.
- (10). No todo lo que se puede saber acerca de las personas es físico.
- (11). El conocimiento físico completo no es conocimiento completo del mundo.

Puesto que para el fisicalismo el mundo es exclusivamente físico, cualquier conocimiento del mismo deberá ser necesariamente físico. Por tanto, un conocimiento físico completo es un conocimiento completo del mundo y viceversa. María ilustra el caso de alguien que aparentemente tenía un conocimiento completo de las personas, que forma parte del conocimiento del mundo, y este era un conocimiento físico. De repente, María alcanza un nuevo conocimiento acerca de las personas (puesto que aceptamos que el saber cómo es algo nuevo es un tipo de nuevo conocimiento acerca de las personas). La conclusión es que, aunque tuviera todo el conocimiento físico de las personas, María no tenía todo el conocimiento de las personas. Puesto que el



conocimiento de las personas forma parte del conocimiento del mundo, y el conocimiento de las personas no se reduce a conocimiento físico, el conocimiento físico del mundo no es conocimiento completo del mundo.

La cuestión de los experimentos mentales

El de María es un caso más de entre una dilatada lista de experimentos mentales, empleados tanto en algunos campos científicos como en filosofía, para ilustrar una determinada posición. La puesta en duda de la validez de estos experimentos como argumentos en contra o a favor de algo podría comprometer, por ejemplo, la dicha crítica que Jackson realiza al fisicalismo. Lo primero que conviene señalar acerca de los experimentos mentales es su naturaleza dispar. Tal y como se presenta en su entrada de la *Stanford Encyclopedia of Philosophy* (Brown & Fehige, 2019), así como por Wilkes en su *Real people: Personal Identity without Thought Experiments* (Wilkes, 2003), la clasificación de estos experimentos es variada. Con todo, sí podemos hallar divisiones con una mayor repercusión que otras. Estas son, por una banda, la división entre los experimentos positivos, aquellos que buscan apoyar una tesis, y los negativos, los que pretenden derrumbar una posición. Por la otra, se sitúa la diferencia entre los experimentos realizables, aquellos que, aunque lejanos son física y lógicamente posibles, y los irrealizables. En el caso de María podríamos dudar acerca de si es realizable que alguien atesore todos los conocimientos físicos relativos a las personas.

A nuestro parecer, el uso de experimentos mentales en filosofía puede resultar relevante en la medida en que ayude a presentar un argumento. Tanto en el caso de los experimentos realizables como en el de los irrealizables. Desde luego, el fructífero papel jugado por estos experimentos en la física, como es el caso del mismo Einstein, parece dar cierto peso a esta opinión. Más allá de esto, la cuestión se sitúa en si el imaginar una hipotética situación nos puede permitir conocer la naturaleza de las cosas o más bien es pura palabrería que nos separa del suelo firme.

Volvamos al ejemplo de María. ¿Qué nos muestra este experimento? Este experimento, con la historia que nos narra, sirve para pensar que no es verdad que el conocimiento de los hechos físicos se equipare al conocimiento del mundo. En el escenario pergeñado por Jackson, se consigue



filtrar el supuesto conocimiento no físico del físico mostrando su existencia. Algo que sin el experimento no sería tan fácil de ver: basta intentar pensar desde nuestra propia posición en la diferencia entre el conocimiento físico y no físico presentada para darnos cuenta de su complejidad. El experimento consigue ilustrar exitosamente un argumento. Pero ahora supongamos que, por el motivo que fuere, la comunidad científica determina que es imposible que alguien consiga poseer todo el conocimiento físico acerca de las personas. Este experimento, al igual que el de las amebas de Parfit (cfr. Parfit, 2004) y otros muchos, resultaría irrealizable. ¿Invalidaría esto al argumento que provee este experimento? No parece disparatado dar una respuesta negativa ya que lo único que mostraría es que María nunca podrá existir, pero la esencia permanecería igual. El experimento mental, sea o no realizable, actúa al modo de un “como si”, de un contrafáctico, para transmitir un argumento apelando a nuestras intuiciones. Podemos recordar, por ejemplo, el *Rebelión en la granja* de Orwell y preguntarnos si el que sea un escenario absolutamente inverosímil es óbice para que esta novela nos sirva para aprender algo nuevo acerca del mundo.

El uso de experimentos mentales creemos que puede resultar útil, pues, en la medida en que permita esclarecer determinadas posiciones. Lo que desde luego no podemos pasar por alto son las diferencias presentes entre los diferentes experimentos. Así como la necesidad de enjuiciarlos y determinar su validez en virtud de su calidad.

Las objeciones de Paul Churchland

En su “What Mary didn’t know” (1986), Jackson recoge tres objeciones presentadas por Paul Churchland (1985). En la primera, Churchland ataca en su esencia el argumento del conocimiento. Acorde al eliminativista, Jackson está hablando en su argumento indistintamente de dos clases de conocimiento. Por una parte, introduce algo así como el “conocimiento por descripción”. Aquel conocimiento que María iría poseyendo mientras se encontraba en la habitación en blanco y negro. El conocimiento físico acerca de las personas. La premisa, según Churchland, es: “María sabe todo lo que hay que saber acerca de los estados cerebrales y sus propiedades” (Jackson, 1986, p. 293). Pero, cuando Jackson nos habla del conocimiento que María adquiere al salir de la habitación, no



está hablando del mismo tipo de conocimiento. Aquí, según Churchland, se nos presenta un “conocimiento por familiaridad”. Estamos, ante una distinción entre el “saber qué” del “conocimiento por descripción”, y el “saber cómo” del “conocimiento por familiaridad”. Este último caso, se presenta cuando Jackson afirma que María aprende algo al percibir por vez primera los colores. El problema para Churchland “es la simplicidad en sí misma” (Jackson, 1986, p. 293). El argumento de Jackson según Churchland se formula como sigue:

(1) María sabe todo lo que hay que saber acerca de los estados cerebrales y sus propiedades.

(2) No es el caso que María sepa todo lo que hay que saber acerca de las sensaciones y sus propiedades.

Por lo tanto, por la Ley de Leibniz,

(3) Las sensaciones y sus propiedades \neq los estados cerebrales y sus propiedades (Jackson, 1986, p. 293).

El argumento muestra la distinción establecida en las premisas. Aproximadamente en esta línea, también irán otros autores como David Lewis y Laurence Nemirow al afirmar que lo que adquiere María es una cierta habilidad. Pero, en ningún caso, un “saber qué”:

Jackson's argument is valid only if 'knows about' is univocal in both premises. But the kind of knowledge addressed in premise 1 seems pretty clearly to be different from the kind of knowledge addressed in (2). Knowledge in (1) seems to be a matter of having mastered a set of sentences or propositions, the kind one finds written in neuroscience texts, whereas knowledge in (2) seems to be a matter of having a representation of redness in some prelinguistic or sublinguistic medium of representation for sensory variables, or to be a matter of being able to make certain sensory discriminations, or something along these lines (Churchland, 1985, p. 23).

En su segunda objeción, Churchland pretende darle la vuelta al argumento de Jackson de modo que este se vuelva en su contra. Busca introducir la posibilidad, en el experimento, de que María



también lo supiera todo acerca del dualismo y de los *qualia*. De este modo, al ser liberada, María también aprendería, al fin y al cabo, algo nuevo. El dualismo dejaría, en virtud de ello, algo por explicar. La tercera objeción apela a la posibilidad de que María hubiese podido acceder a ese “saber cómo” de los colores mediante la imaginación. No podemos menospreciar el hecho de que María aglutina todo el bagaje de conocimiento físico posible. Quizás, sin haber nunca visto el rojo de los tomates, María es capaz de hacerse una idea del mismo.

Pero retornemos a la primera objeción de Churchland. Este afirma, como veíamos, que el argumento de Jackson tan sólo podría llegar a buen puerto si el conocimiento al que se refiriera en ambas premisas fuese el mismo. Mas parece que deberíamos distinguir el conocimiento entendido como un conjunto de enunciados que uno podría encontrar en textos de neurofisiología, de las diferentes representaciones que nos podamos hacer. No es lo mismo decir que el rojo es el color que ocupa el primer lugar en el espectro luminoso y que posee una determinada longitud de onda, que observar el rojo de la sangre. Y, sin embargo, parece que Jackson equipara ambos tipos de fenómenos al afirmar que a María algo le debía de quedar en el tintero cuando, a pesar de poseer todo el conocimiento neurofisiológico, al ver el rojo aprende algo nuevo. El *non sequitur* del que peca Jackson se mostraría con meridiana claridad visto así su argumento:

- (a) Mary has mastered the complete set of true propositions about people's brain states.
- (b) Mary does not have a representation of redness in her prelinguistic medium of representation for sensory variables.

Therefore, by Leibniz's law, (c) The redness sensation \neq any brain state (Churchland, 1985, p. 24).

Jackson considera esta objeción de Churchland. Afirma aquel que, si lo expuesto por Churchland fuera su argumento, él mismo estaría de acuerdo en su implausibilidad. Pero este no es el caso. De hecho, en su argumento, dice Jackson, “lo que directamente viene al caso” no son los tipos de conocimiento de los que podamos hablar, sino “lo que sabe” María. Lo que sí considera Jackson es la posibilidad de que la premisa que sostiene que María aprende algo nuevo al ser



liberada no sea cierta. Si no es cierto que María aprende algo nuevo al percibir los colores, entonces todo el argumento se derrumba. ¿Puede que ese algo que María adquiere al ser liberada no sea propiamente un conocimiento sino una especie de habilidad? Jackson está de acuerdo en que cuando ve por primera vez un tomate, María comienza a adquirir una serie de habilidades que antes no tenía. La capacidad de imaginarse el color rojo, aunque no lo tenga delante, o la capacidad de recordar el rojo del tomate que vio, son algunos ejemplos. Sin embargo, a Jackson esto le sabe a poco. A modo de intuición, Jackson nos propone imaginarnos que, en su excepcionalidad, María también recibió lecciones acerca de escepticismo; concretamente, acerca de las otras mentes. Así, tras ver los colores, María duda acerca de si en verdad puede decir que sabe algo más de las personas. Acerca de si acababa de obtener o no conocimiento fáctico acerca de los demás. Desde luego, dice Jackson, esto debe de ser así pues no le parece que María haya podido “romperse la cabeza” acerca de las habilidades de los demás. La habilidad fue “una constante conocida en todo momento”, entonces, de lo que debe dudar es del susodicho conocimiento del que carecía. Más allá de este incierto ejemplo, Jackson reconocerá que el suyo es un argumento *a priori*, de cuyas premisas no tiene pruebas. Pero, en cualquier caso, premisas “altamente plausibles” de un “argumento válido”.

¿Qué podría decir de nuevo Churchland acerca de la respuesta de Jackson? Tengamos presente el argumento que con el que Jackson afirmó corregir la interpretación de Churchland:

(1) María (antes de ser liberada) sabe todo lo físico que hay que saber acerca de otras personas.

(2) María (antes de ser liberada) no sabe todo lo que hay que saber acerca de otras personas (porque ella aprende algo acerca de ellos al ser liberada).

Por consiguiente:

(3) Existen verdades acerca de los demás (y de ella misma) que escapan a la explicación fisicalista (Jackson, 1986, p. 115).



Churchland podría aceptar que el argumento está aquí mejor explicitado que de la forma en que él lo había hecho. También podría estar de acuerdo en que este argumento es válido. El meollo del asunto se sitúa en la verdad de las premisas. El problema al que Churchland hacía referencia en su primera objeción todavía continúa perviviendo. De la primera premisa, aceptando el experimento de María, no hay nada que decir. Se puede suponer con Jackson que María tiene todo el conocimiento físico, que podríamos denominar “proposicional”, referente a las personas. A cualquier pregunta de esta índole, María puede dar una respuesta. No sucede lo mismo con la segunda premisa. Jackson persiste en el error de confundir dos tipos de conocimiento diferente: el conocimiento que podríamos decir “proposicional” o “descriptivo” y el “conocimiento por familiaridad”, propio de la percepción. Ambos tipos se refieren a lo mismo, por ejemplo, al color rojo (sea cual sea la variedad), pero la gran diferencia está en el modo de conocer, no en la naturaleza de lo conocido. El argumento de Jackson sigue sin llegar a buen puerto al pretender romper la barrera existente entre los modos de conocer y el conocimiento mismo. María no aprende nada nuevo al salir de la habitación, sino que adquiere una representación de lo que ya sabía.

Referencias

- Arias, A. (2011) “¿Ha logrado Dennett quinea los *qualia*? Una revisión naturalista”, *A Parte Rei. Revista de filosofía*, (73), 1-26.
- Arias, A. (2021) *Introducción a la ciencia de la conciencia: El estudio de la experiencia subjetiva en filosofía, psicología y neurociencias*, Madrid: Los libros de la catarata.
- Brown, J. & Fehige, Y. (2019) “Thought Experiments”, en Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/thought-experiment/>
- Byrne, A. (2020) “Inverted Qualia”, en Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/qualia-inverted/>
- Chalmers, D. (1999) *La mente consciente: En busca de una teoría fundamental*, Barcelona: Gedisa.



- Churchland, P. M. (1985) "Reduction, Qualia, and the Direct Introspection of Brain States", *The Journal of Philosophy*, (82), 8-28.
- Dennett, D. (1995) *La conciencia explicada: Una teoría interdisciplinar*, Barcelona: Paidós.
- Dennett, D. (2003) "Quinear los *qualia*", en Maite Ezcurdia y Olbeth Hansberg (comp.) *La naturaleza de la experiencia, vol. I: Sensaciones*, México: UNAM, 213-262.
- Jackson, F. (1982) "Epiphenomenal Qualia", *Philosophical Quarterly*, (32), 127-136.
- Jackson, F. (1986) "What Mary Didn't Know", *The Journal of Philosophy*, (83), 291-295.
- Levine, J. (1983) "Materialism and Qualia: The Explanatory Gap", *Pacific Philosophical Quarterly*, (64), 354-361.
- Nagel, T. (1974) "What Is it Like to Be a Bat?", *The Philosophical Review*, (83), 435-450.
- Wilkes, K. (2003) *Real People: Personal Identity without Thought Experiments*, Oxford: Oxford University Press.
- Parfit, D. (2004) *Razones y personas*, Madrid: Antonio Machado Libros.