

La filosofía de la mente de Thomas Metzinger: neurofenomenología, consciencia artificial y pesimismo



Thomas Metzinger's philosophy of mind: neurophenomenology, artificial consciousness and pessimism

JOSÉ CURBERA-LUIS

Universidad de Salamanca (España)

Fecha de envío: 29-06-2022

Fecha de aceptación: 24-10-2022

DOI: 10.24310/Claridadescrf.v15i2.15061

RESUMEN

En el presente artículo hacemos una revisión del método neurofenomenológico de Thomas Metzinger y su capacidad para responder a problemas habituales de filosofía de la mente y evitar dificultades de otros materialismos. Metzinger define estados mentales en función de estados cerebrales, sin embargo, frente al materialismo eliminativo, sí considera como real la experiencia fenoménica y

rechaza su reducción conceptual o epistémica. Esta peculiaridad salva su propuesta de críticas como la de José Ignacio Murillo o Graham Harman, que parecen dirigidas a otros tipos de materialismos. Por ello, consideramos necesario repensar las consecuencias éticas de la tajante negación del libre albedrío y la afirmación de la posibilidad de máquinas artificiales conscientes que conlleva la propuesta de Metzinger.

Claridades. Revista de filosofía 15/2 (2023), pp. 75-101.

ISSN: 1889-6855 ISSN-e: 1989-3787 DL.: PM 1131-2009

Asociación para la promoción de la Filosofía y la Cultura en Málaga (FICUM)

PALABRAS CLAVES

Thomas Metzinger; materialismo, neurofenomenología; automodelo transparente; túnel del yo

ABSTRACT

In this article we review Thomas Metzinger's neurophenomenological method and its ability to respond to some problems in philosophy of mind, avoiding difficulties present in other kinds of materialism. Metzinger defines mental states according to brain states, however, unlike eliminative materialism, he does consider phenomenal experience as real and rejects its

conceptual or epistemic reduction. This peculiarity saves his proposal from criticisms such as that of José Ignacio Murillo or Graham Harman, which seem directed at other types of materialisms. Therefore, we consider it necessary to rethink the ethical consequences of the categorical denial of free will and the affirmation of the possibility of conscious artificial machines that Metzinger's proposal entails.

KEYWORDS

Thomas Metzinger; materialism; neurophenomenology; transparent self-model; ego tunnel

1. INTRODUCCIÓN: LA NEUROFENOMENOLOGÍA DE METZINGER

Metzinger propone un método neurofenomenológico donde se definen estados fenoménicos en virtud de estados cerebrales. Metzinger considera que la mente es un proceso físico, la mente es ontológicamente reductible al cerebro. Sin embargo, frente al materialismo eliminativo, Metzinger sí considera como ontológicamente real la experiencia fenoménica en primera persona y no considera que sea epistemológicamente reductible ni eliminable. A partir de esta metodología concluye que la condición que hace que surja la experiencia fenoménica y el «yo» es que se genere un automodelo fenoménico transparente.

El filósofo de la mente Thomas Metzinger se propone al principio de *El túnel del yo: ciencia de la mente y mito del sujeto* un proyecto ambicioso:

convencer al lector de que no existe, más bien, de que no existe tal cosa como su «yo». El planteamiento que hace de este proyecto empieza de forma similar a otros proyectos materialistas: con una orientación analítica y dándole gran importancia a los datos aportados por la neurociencia (Metzinger, 2018: 12). Sin embargo, pronto intenta distinguirse de otros autores como Dennett, pues les achaca darle poca importancia a los datos aportados por la fenomenología con sus propuestas eliminativistas.

En vez de disolver como Dennett conceptos como «ser un yo» o «subjetividad», experiencias directas indudables como sentir que una mano es mía o ver la rojez o preguntas como «quién está leyendo estas líneas» tratándolas de «psicología popular», afirma que el sentido de identidad y de interioridad son mucho más profundos que un mero «estar en un mundo simulado en el cerebro» y que este va a ser el principal tema que va a tratar. Ante nosotros aparece un mundo, una consciencia que «es mucho más que física sumada a biología» y «hace que la realidad se aparezca dentro de sí» (Metzinger, 2018: 23-29).

La respuesta que va a dar Metzinger la introduce al principio del libro, sin embargo, la desarrolla posteriormente ejemplificándola en diversos experimentos neurológicos que buscan lograr una forma de «neurofenomenología» (Metzinger, 2003: 83). A partir del estudio de sueños lúcidos, experiencias extracorporales, miembros fantasmas, etc., la conclusión que parece seguirse es que el rasgo necesario y suficiente para «ser un yo» es tener un automodelo fenoménico, una «identidad-como-corporeización». Este automodelo es un modelo generado por procesos neuronales que incorpora todos los rasgos fenoménicos de ser un yo en un mundo externo en un momento determinado. Metzinger lo define como: «un único modelo de realidad coherente y temporalmente estable que se centra representacionalmente alrededor o sobre un único sujeto fenoménico coherente y temporalmente estable» (Metzinger, 2003: 159). Es decir, a raíz de este conjunto de correlatos neuronales se produce la «aparición de un mundo». Estos procesos neuronales simulan la existencia del cuerpo en relación con el mundo externo y fenoménicamente nos sentimos identificados con ese cuerpo; se genera así el «yo».

Se llega a esta conclusión a través del método neurofenomenológico: el estudio comparado de diversos casos de alucinaciones, sueños lúcidos y no lúcidos, miembros fantasma, enfermedades mentales, «muertes del

ego», experiencias extracorporales, etc. A través de estos experimentos, Metzinger concluye que la experiencia fenoménica de la intencionalidad o de la funcionalidad no son necesarias para la aparición del sujeto; sino que son propiedades «añadidas» que no están presentes en sueños, en ciertas experiencias extracorporales o determinadas enfermedades mentales, en las cuales sí que existe la experiencia fenoménica del «yo», ligada al automodelo fenoménico. Volveremos a la discusión de estos casos cuando tratemos el concepto de *qualia* en Metzinger y, evidentemente, cuando discutamos su noción de libre albedrío.

El automodelo fenoménico es un proceso que Metzinger llama «transparente» (Metzinger, 2018: 61-70), es decir, no podemos ver cómo se produce. El autor nos insta a imaginarnos los procesos mediante los cuales se genera la experiencia fenoménica holística (vista, tacto, olor...) de tener entre nuestras manos su libro y leerlo. Evidentemente es imposible la reducción por parte del sujeto de esta experiencia fenoménica a los procesos que la generan (sean cuales sean), pues, epistemológicamente, lo más inmediato para nosotros es esa misma experiencia. A partir de la primera persona, cualquier proceso que supuestamente genere la experiencia fenoménica nos es completamente invisible. Esto es lo que llama Metzinger el «túnel del yo»; recorreremos un túnel de experiencias fenoménicas cuyas paredes nos son completamente invisibles, en primera persona es imposible saber cómo lo estamos recorriendo. Sin embargo, para Metzinger esta metáfora se rompe, ya que no hay nadie recorriendo el túnel; no hay nada más que el propio túnel de experiencias fenoménicas. A pesar de esto, como estas experiencias fenoménicas están insertadas en un automodelo que lleva consigo la experiencia de la «auto-identificación», con el contenido del cuerpo consciente, y «auto-localización», en un marco espacial, nos parece que hay «alguien» que recorre el túnel (Metzinger, 2018: 150-152).

Además, es un automodelo bastante limitado, incluso errado, ya que los pocos aspectos de la realidad con los que nos pone en contacto siempre son los de hace unas fracciones de segundo. Esto se debe a que ese es el tiempo que tardan las neuronas en generar el automodelo. También se producen diversos errores en nuestro automodelo, como cuando se siente un miembro fantasma como si todavía fuera parte de nuestro cuerpo o

cuando incorporamos a nuestro automodelo una mano de goma, que sustituye a la real.

Por ello, Metzinger nos llama «realistas ingenuos». Nuestro automodelo fenoménico nos presenta fenoménicamente en contacto directo con unos aspectos determinados de la realidad. Los procesos que generan la «aparición del mundo» son subjetivamente invisibles, al igual que los demás aspectos de la realidad «externa» que no pertenecen a nuestro automodelo. De hecho, desde la perspectiva reduccionista que toma, el contacto con la realidad es de cierta manera ilusorio, ya que todos estos procesos se están generando en el espacio cerrado de nuestro cuerpo. Esta idea ya está presente en Spinoza, a quien cita el propio Metzinger: las ideas que tenemos de los cuerpos exteriores (que para Metzinger serían este automodelo fenoménico) indican más de la constitución de nuestro cuerpo que de los objetos exteriores (Metzinger, 2018: 157).

El proyecto que presenta Metzinger parece relativamente sencillo: se busca correlacionar la experiencia mínima de «ser un yo» con un correlato neuronal. Después, se identifica la experiencia fenoménica de la intencionalidad y del libre albedrío con sus respectivos correlatos. Tras esto, se experimenta con estos circuitos neuronales para mostrar casos donde claramente no se está tomando una decisión libre pero sí aparece ese correlato neuronal y la respectiva experiencia fenoménica.

2. «PRINCIPIO DE INEFABILIDAD», *QUALIA* Y REDUCCIÓN

Thomas Metzinger, antes de aventurarse a articular las connotaciones filosóficas de los descubrimientos científicos, tiene que discutir los problemas que posee el intento de naturalización de la mente. Todo el discurso previo, sin una caracterización ontológica de la mente ni de los procesos mentales, no es más que una justificación circular. Es crucial en la justificación de su materialismo y su metodología neurofenomenológica la discusión que hace del concepto de *qualia*. La definición de *qualia* que va a discutir Metzinger es:

Forma especial de contenido mental, de la cual es cierto que:

1. Se dispone de criterios de identidad subjetivos, mediante los cuales podemos reconocer introspectivamente su identidad transtemporal;

2. Es una forma máximamente simple y experimentalmente concreta (es decir, máximamente determinada) de contenido, sin características estructurales internas;
3. Provoca la instanciación de una propiedad no física de primer orden, una propiedad fenoménica
4. No existe un mapeo sistemático uno a uno de esas propiedades subjetivas a propiedades objetivas;
5. Se capta de forma directa, intuitiva y epistémicamente inmediata;
6. Es subjetivo al ser captado «desde la perspectiva de la primera persona»;
7. Posee un núcleo fenoménico intrínseco, que, analíticamente, no puede disolverse en una red de relaciones; y
8. Los juicios sobre esta forma de contenido mental no pueden ser falsos (Metzinger, 2003: 93).

En el primer punto recoge que podemos identificar introspectivamente, privadamente, el mismo contenido mental a lo largo del tiempo; en el segundo se expresa que es una experiencia simple, no compuesta por partes; en el tercero se afirma que dan lugar a una propiedad fenoménica, que es no física; en el cuarto se niega la posibilidad de mapearlas con propiedades objetivas; en el quinto se lleva la atención a su inmediatez; en sexto, a su carácter subjetivo; en el séptimo se afirma que no se puede disolver en una red de relaciones, haciendo referencia a su carácter simple y a la imposibilidad de explicar los *qualia* a través de relaciones (presumiblemente, entre neuronas); y en el octavo se afirma que los juicios sobre *qualia* no pueden ser falsos. Esta definición se corresponde con la dada por Lewis; Metzinger reconoce que muchos filósofos se opondrán a esta definición, sin embargo, la discusión va a girar en torno a los dos primeros puntos, que considera que son suficientes (Metzinger, 2003: 93).

Metzinger va a argumentar que los *qualia*, en este sentido tradicional que impide cualquier posible reducción, no pueden existir; sin embargo, no a través de una propuesta eliminativa (no al menos de todos los tipos de *qualia*), sino que propondrá otra forma de conceptualizar los *qualia*.

Metzinger afirma que se debe tener en cuenta el problema de los *qualia* y no simplemente tratar de eliminar el problema siguiendo una metodología lingüística-analítica (Metzinger, 2003: 70-71).

Metzinger afirma que se pueden entender los estados mentales a través de tres propiedades: el contenido representacional o intencional, su relación causal con otros estados internos y por su contenido fenomenal o experimental. Metzinger se propone buscar cuál es la forma más simple de contenido mental, el que encajaría con la definición de *qualia* para posteriormente analizarlo (por ejemplo, la intencionalidad no estaría presente en sueños o alucinaciones, no es necesaria para tener una experiencia fenoménica. Metzinger se opondría, junto a Heidegger, a Husserl). Sin embargo, a través del análisis fenomenológico, Metzinger no encuentra ningún criterio transtemporal para comparar dos experiencias fenoménicas. ¿A qué se debe eso? A la imposibilidad de lograr conceptos de nuestros estados mentales privados utilizando exclusivamente la memoria (Metzinger, 2003: 71-73).

A esto, Metzinger lo llama, el problema de la inefabilidad. Tenemos unos esquemas generales acerca del color, pero las diferencias entre dos tonos de verde («verde no. 24» y «verde no. 25») son completamente inefables para un sujeto una vez ha pasado la experiencia; las diferencias son demasiado sutiles y no pasan más allá de una conceptualización general de «verde claro». La idea de *qualia* como átomos de experiencia irreductibles y distintos parece desvanecerse en favor de una idea de la experiencia más fluida, se puede eliminar el concepto de «átomo de la experiencia». Sin embargo, esto es un problema para el proyecto naturalista, ya que la «inefabilidad» no obra en su favor. Sin la posibilidad de que el sujeto se forme fenoménicamente conceptos que se pueden expresar (para que estas experiencias fenoménicas no sean como los escarabajos de Wittgenstein), parece imposible un proyecto que permita correlacionar los estados subjetivos con los estados objetivos neuronales que lo producen.

Dados estos pronósticos, parece que habría que adoptar un materialismo eliminativo como el de Churchland: se niega la existencia de los *qualia*, se afirma que esas teorías fenoménicas están erradas epistémica y ontológicamente y nos conduciría a un estado que Metzinger considera de ciencia ficción, donde no deberíamos comunicar nuestras experiencias con términos como «verde» sino con jerga neurobiológica. Metzinger rechaza

radicalmente el materialismo eliminativo. Afirma con total rotundidad la realidad e inherente subjetividad de las experiencias fenoménicas y su irreductibilidad epistémica. Aunque, poco después, afirma también su reductibilidad ontológica, hay un único proceso físico conocido que se puede conocer en primera y en tercera persona (Metzinger, 2018: 72-88).

En esta exposición nos encontramos enormes connotaciones ontológicas. La primera es que se ha afirmado con total seguridad el materialismo; Metzinger se opone a los naturalistas como Searle que afirman que la consciencia es ontológicamente distinta del cerebro. Esta visión tiene ciertas ventajas: Metzinger no tiene que explicar la causalidad entre dos planos ontológicamente distintos (la causalidad arriba-abajo no es una explicación suficiente para explicar cómo interactúan dos sustancias ontológicamente distintas), tampoco se enfrenta a la objeción que se le hace a los epifenomenalistas de que si la consciencia es algo causalmente inactivo, simplemente añadida. Para Metzinger los procesos físicos que generan experiencias fenoménicas *son* las experiencias fenoménicas conocidas desde otra perspectiva; no sería posible el *zombie* de Chalmers. A través de un proceso evolutivo se generan primero automodelos inconscientes (procesos físicos sin experiencia fenoménica) que se van complicando gradualmente hasta generar gradualmente una experiencia fenoménica. Para Metzinger este es un proceso gradual donde, a partir de estos automodelos exclusivamente objetivos, se generan automodelos que conllevan una experiencia subjetiva, la cual tiene una ventaja evolutiva (Metzinger, 2018: 87-88).

La gradualidad en la aparición de la consciencia le lleva a tener que postular un doble aspecto; la materia tendría una posibilidad de conocerse en primera persona, pero solo a partir de la generación de un automodelo fenoménico. Metzinger no es pampsiquista, lo mental no es intrínseco a lo material; solo se puede hablar de lo mental una vez lo físico genera un automodelo fenoménico. De esto también se deduce que es perfectamente posible la existencia de *robots* conscientes; se aleja del criterio cognitivista basado en la inteligencia y la capacidad para resolver problemas, hacerse pasar por un humano, etc. El criterio ahora va a ser que produzca un automodelo fenoménico.

Una vez comentadas las consecuencias ontológicas a la solución que nos da Metzinger del «problema de la infabilidad», pasamos a comentar cómo

reduce las experiencias fenoménicas a los hechos materiales, teniendo en cuenta que previamente había rechazado que fueran epistemológicamente reductibles. Este problema lo discute a través de dos casos:

a. Los conceptos en un lenguaje público pueden referirse al contenido fenoménico de estados mentales: por ejemplo, al contenido fenoménico simple (...). A nivel de objeto, los criterios de identidad lógica al usar tales expresiones son experiencias introspectivas, por ejemplo, la experiencia subjetiva de igualdad (...). La psicología o algunos tipos de fenomenología filosófica proporcionan ejemplos de tales lenguajes.

b. Los conceptos en un lenguaje público pueden referirse al contenido fenoménico de estados mentales: por ejemplo, a contenido fenoménico simple. A nivel metalingüístico, los criterios de identidad lógica aplicados cuando se utilizan tales conceptos son propiedades de acceso público, por ejemplo, los del correlato neural de este contenido sensorial activo, o algunas de sus propiedades funcionales (Metzinger, 2003: 82).

Para Metzinger, el primer tipo de fenomenología es imposible debido a la imposibilidad de mantener con seguridad la identidad del contenido fenoménico usando un criterio introspectivo; mientras que el segundo, la «neurofenomenología», es el método de estudio por el que aboga y mediante el cual llega a sus conclusiones. Es imposible que ciertos tipos de fenomenología constituyan un lenguaje público, ya que se quedan en un lenguaje privado sin criterios para mantener constante el significado de sus conceptos.

En la obra de Metzinger se reconoce la imposibilidad de reducir epistemológicamente los estados mentales a estados neuronales. Sin embargo, metodológicamente, se definen los primeros en virtud de los segundos, para asegurarse de que el criterio de significación es público y no privado. A través de esta metodología, es posible discutir los conceptos de «intencionalidad» y «funcionalidad»; para, posteriormente, buscar el principio de *qualia* mínimo con un criterio distinto:

Formas de contenido fenoménico máximamente simple, determinado y desambiguado. Es una suposición empíricamente plausible que los criterios de identidad transtemporales, así como lógicos, se pueden desarrollar desde una perspectiva en tercera persona, mediante la investigación de las propiedades de los correlatos físicos mínimamente suficientes de contenido sensorial simple, a los que se puede acceder mediante la investigación neurocientífica (...) o mediante el análisis funcional (...) del papel causal realizado por estos

correlatos. (...) Las formas más simples de contenido fenomenal pueden individualizarse funcionalmente (Metzinger, 2003: 84).

Como se puede ver, este es el colofón evidente de lo discutido previamente. Ya se ha establecido la metodología que se va a seguir para analizar diversas experiencias fenoménicas, los «estudios de casos neurofenomenológicos» van a ocupar el grueso de *Being No One*. La conclusión a la que llegará ya se ha reiterado repetidamente: la forma de contenido fenoménico máximamente simple, determinado y desambiguado es el automodelo fenoménico; al cual se le añadirá posteriormente la intencionalidad, el arbitrio, etc.

3. SOBRE MANOS DE GOMA, MIEMBROS FANTASMAS Y EXPERIENCIAS EXTRACORPORALES

A continuación, vamos a exponer alguno de los experimentos que pone de ejemplo Metzinger en la segunda parte de *El túnel del yo*, donde se ponen ejemplos de experimentos realizados con la metodología antes propuesta. Las consecuencias de estos experimentos son perfectamente predecibles una vez se ha expuesto su ontología, a pesar de esto, vamos a tratarlos brevemente para comprender la concepción antropológica a la que llega Metzinger. Es interesante anotar que toda la ontología de Metzinger la hemos discutido sin apenas entrar en los experimentos, lo cual refuerza la crítica que señala su aprioridad. Esta exposición también nos servirá como antecedente a la ética que propone el autor.

Metzinger comienza exponiendo sus experimentos fundamentales, aquellos que muestran más claramente que el «yo» no es más que un automodelo fenoménico. Metzinger nos describe un experimento en el cual se oculta la mano real del sujeto mientras que se le pone a la vista una mano de goma en una posición semejante, tras cierto tiempo estimulando tanto la mano real como la de goma, la de goma —que es la única visible— pasa a ser parte del automodelo del sujeto *sustituyendo* a la mano real, no añadiéndose a ella (Metzinger, 2018: 101-111).

Algo semejante ocurre con los miembros fantasma. Metzinger cuenta el caso de diversos pacientes con miembros fantasmas dolorosos y que los sujetos «notan» inmóviles. A través de una simple caja y de un espejo,

el paciente puede ver el reflejo de su mano en el lugar de su miembro fantasma. Si mueve ambas manos simultáneamente, el miembro fantasma se incorpora al automodelo y puede ejercer su voluntad sobre el mismo. En cuanto se retira el espejo, el miembro fantasma vuelve a estar inmóvil (Metzinger, 2018: 150-155).

Metzinger también experimenta con las experiencias extracorporales. En estas experiencias extracorporales, el sujeto siente que se sale de su propio cuerpo, lo deja atrás y lo continúa viendo; en ocasiones es capaz de flotar, explorar la habitación, incluso salir por la ventana. La fenomenología de las experiencias extracorporales parece conducir al dualismo y a la existencia de un «cuerpo sutil», inmaterial, pero extenso. Metzinger considera que estas experiencias del «cuerpo sutil» abandonando el cuerpo pueden haber dado origen a las teorías acerca del «soplo de vida», como el *spiritus* latino y el *pneuma* griego, que son antecedentes del concepto de «alma» (Metzinger, 2018: 112-117).

Las experiencias extracorporales son realistas, pero solo parcialmente. El cuerpo percibido en tercera persona es distinto a cómo debería ser en realidad (tamaño, ropa, peinado, etc.), el cuerpo flotante en ocasiones arroja sombra y en otras ocasiones no, etc. La explicación que da Metzinger de las mismas es que se están separando dos elementos de nuestro automodelo: el yo corpóreo y nuestra visión (que generan, respectivamente, nuestra autocorporeización y nuestra autoidentificación). Estas dos dimensiones son las que dan origen a nuestro automodelo, pero en ciertos casos pueden separarse. El *locus* de la identidad se localiza en la visión; lo cual es perfectamente coherente respecto a los experimentos de la mano de goma y del miembro fantasma, ya que en estos experimentos se veía como el *input* visual permitía cambiar el automodelo (Metzinger, 2018: 117-137).

Esta desconexión entre las partes de nuestro automodelo y la ausencia de *input* sensorial hace que el cerebro tenga que recurrir a fuentes de información internas. La psicóloga Blackmore nos insta a recordar, por ejemplo, un día que pasásemos en la playa. Parte de estos recuerdos no estarán codificados en primera persona, sino en vista de pájaro simulando nuestro cuerpo paseando por la costa. Esto explicaría la capacidad del cerebro en las situaciones extracorporales para simular la apariencia física del cuerpo y que, en muchas ocasiones,

sea ligeramente errónea, lo cual no tendría sentido si no fuera una simulación del cerebro (Metzinger, 2018: 118). A través del tratamiento de las experiencias extracorporales, Metzinger ha conseguido dar una explicación naturalista a un fenómeno que antes era considerado pseudocientífico o un argumento a favor del dualismo. Además, ha conseguido dar ejemplos experimentales de su teoría del automodelo fenoménico.

Estos experimentos apuntan a que la experiencia fenoménica es una experiencia holística que integra todos nuestros sentidos con el fin de lograr una imagen de dónde está nuestro cuerpo lo más acertada posible:

Lo que se siente en la ilusión de la mano de goma, [lo que se siente con los miembros fantasma] es exactamente lo mismo que siente usted cuando atiende a la sensación de sostener este libro entre las manos, ahora mismo. (...) El trabajo del cerebro es simular el cuerpo *para el cuerpo*. (...) Exclusivamente en el plano de la experiencia consciente, usted vive su vida en un cuerpo virtual y no uno real (Metzinger, 2018: 158).

En estas declaraciones se puede ver el tono ominoso de gran parte de las conclusiones de Metzinger (como que «no somos nadie»), en este caso, este concepto de que vivimos en una simulación, en una realidad virtual, viene del simple hecho de que nuestro contacto con la realidad está mediado por procesos cognitivos que son transparentes para nosotros y, en ocasiones, falibles. Estas ideas no son especialmente nuevas ni ajenas a la tradición filosófica y científica; sin embargo, es interesante que desde una perspectiva materialista se busque explicar (y no eliminar) experiencias fenoménicas tan poco habituales e interesantes como pueden ser las experiencias extracorporales.

Además, la teoría del automodelo nos da unas herramientas para entender estas experiencias a través de los conceptos de «auto-corporeización» y «auto-identificación»; que son procesos fenoménicos generados por correlatos neuronales que no necesitan la existencia de este cuerpo simulado; ya sea porque no forma parte de nuestro cuerpo (manos de goma o miembros fantasmas) o porque todo este proceso se hace a través de fuentes de información internas, sin *input* sensorial (en las experiencias extracorporales). Es decir, se produce un desligamiento de la experiencia fenoménica de nuestra corporeidad real y pasa a estar inscrita en la corporeidad simulada por estos correlatos neuronales. Lo

que es especialmente interesante de las experiencias extracorporales es que demuestran que para que tenga lugar una experiencia fenoménica no es una condición necesaria que coincida nuestro *locus* de la identidad con la corporeización del cuerpo simulado.

4. ALGUNAS CRÍTICAS INSUFICIENTES

Al contrario que los materialistas eliminativistas, sí que se le está dando importancia a la experiencia fenoménica, aunque se ha reducido *metodológicamente* a estados neuronales. La ontología de los estados fenoménicos no es una conclusión a la que pueda llegar aplicando esta metodología sin caer en una petición de principio, pues esta ontología es el supuesto previo sobre el que se levanta el método neurofenomenológico que usa para llegar a sus conclusiones. Por ello, algunos autores han acusado a Metzinger de dar por supuesto lo que se trata de demostrar (Arana, 2015: 144).

Sin embargo, esta afirmación no es del todo exacta porque la justificación de su ontología y metodología está separada del uso de su metodología para llegar a otras de sus conclusiones. Por ejemplo, una de sus justificaciones para no tener que postular otra entidad ontológica es, simplemente, que puede explicar el concepto de «sujeto de experiencia» a través del concepto objetivo de «sistema que genera un automodelo» sin necesidad de postular susodichas entidades. Metzinger no ve necesario postular un aspecto de la realidad que se corresponda con la forma de entender la mente en la psicología popular o la fenomenología; le basta con «sistemas operando bajo automodelos transparentes», la subjetividad no es nada ontológicamente distinto, sino una propiedad de estos sistemas (Metzinger, 2003: 576-577).

Una de las críticas realizadas a Metzinger es por parte de José Ignacio Murillo, de la Universidad de Navarra. Murillo hace un breve comentario sobre la teoría de la verdad que escoge Metzinger, la cual identifica con una teoría pragmatista de la verdad para escapar del regreso al infinito que provocan las teorías representacionistas de la verdad (Murillo, 2011: 121).

Posteriormente afirma que los modelos fenoménicos de Metzinger se corresponden con los fantasmas aristotélicos, las imágenes que producen las alucinaciones. Luego afirma:

(...) la misma capacidad de distinguir una alucinación de una imagen adecuada presupone una forma de conocimiento distinto: la inteligencia, que, a diferencia de los sentidos, objetiva y conoce en términos de realidad. Si aceptamos esta salvedad, carece de sentido afirmar que la imagen del mundo que actualmente consideramos no se corresponde con lo que está ocurriendo fuera del mundo. (...E)sas imágenes no son en sí mismas una ocultación de la realidad, sino un modo de conocerla. El procesamiento cerebral que queda oculto a la hora de conocer no es conocimiento, sino, precisamente lo que lo hace posible (...) El conocimiento sensible se da gracias a procesos físicos y orgánicos, pero no se identifica con su ocurrir. Conocer los procesos que hacen posible el conocimiento sensible sería conocer algo distinto de lo que se conoce gracias a ellos, pero en ningún caso se identifica con un proceso cerebral (Murillo, 2011: 122).

Hemos incorporado un fragmento relativamente extenso para caracterizar lo mejor posible la opinión de Murillo, que se podría resumir en que los modelos fenoménicos no ocultan la realidad, sino que son un modo de conocerla y que conocer los procesos físicos y orgánicos no es lo mismo que conocer lo que se conoce a través de ellos. Como hemos mencionado, Metzinger rechaza el reduccionismo epistemológico, así que no tiene sentido que la última afirmación de Murillo pretenda ser una crítica, en palabras de Metzinger:

A pesar de que la conciencia es un proceso físico, estas dos formas de saber [primera y tercera persona] no pueden ser integradas. Conocer hasta las últimas partes del cerebro de una persona nunca autorizará a saber nada de la persona como tal (Metzinger, 2018: 88).

La primera crítica también la contempla Metzinger: «aunque hubiéramos nacido con la capacidad de percibir nuestro procesamiento interno, seguiríamos conociendo la realidad bajo la forma de la representación» (Metzinger, 2018: 69). Metzinger contempla estas representaciones como formas de conocer la realidad, las cuales permiten descubrir que los procesos mediante los cuales se generan son opacos y no se pueden descubrir subjetivamente; ahí reside el «ocultamiento» o la «ilusión». No se aleja excesivamente de la visión de Spinoza del «alma como idea del cuerpo». Metzinger también menciona las alucinaciones en las cuales se es consciente de que se está alucinando como momentos en los cuales una

parte del «túnel del yo» se vuelve opaco, es decir, es posible darse cuenta de que parte de nuestro contenido fenoménico es una representación generada por nuestro propio cerebro (Metzinger, 2018: 68).

Murillo también se pregunta, citando a Husserl, ¿es posible esta teoría si es verdad lo que afirma? Las conclusiones de Metzinger niegan el sujeto, pero presuponen la posibilidad del conocimiento filosófico y científico, que requiere de un sujeto. Metzinger comenta al respecto:

Parece haber dos alternativas:

O lo ve como una descripción real de un conjunto de posibilidades que pueden ser nomológicamente probable (es decir, empíricamente plausible) y conceptualmente coherente (es decir, filosóficamente plausible) al mismo tiempo. Entonces *no se puede convencer*. Llame a esto la «alternativa científica del dilema.» No puede estar convencido, porque la idea de que no existen cosas tales como el yo — incluido el suyo — en el mundo sigue siendo estrictamente contradictorio, una imposibilidad fenomenal. Ahora puede recurrir a la otra alternativa. Llamemos a esto la «alternativa espiritual del dilema». Podría cambiar su modelo fenomenal global del mundo de una manera que lo convierte en una posibilidad. (...) Su nueva configuración neurofenomenológica correspondería entonces a lo que anteriormente se denominó «conciencia del sistema», es decir, un estado fenoménicamente no subjetivo de la conciencia. En este caso, no podría formar verazmente las oraciones «Yo» correspondientes. Y, por lo tanto, ni siquiera podría autoatribuirse su nueva configuración neurofenomenológica a sí mismo. En conclusión, no se puede estar convencido de la verdad de la teoría del automodelo de la subjetividad, necesariamente (Metzinger, 2003: 642-643).

Básicamente, una vez se ha presentado esta teoría hay dos opciones, afirmar «Yo no soy un sujeto», lo cual es una contradicción; o cambiar mi modelo fenoménico, lo cual es imposible, para poder afirmar que, de hecho, no se es un sujeto. Sin embargo, no se puede afirmar que no se es un sujeto, porque un no-sujeto no puede afirmar nada del tipo «Yo soy X». Ninguna de las alternativas nos permite afirmar la veracidad de la teoría de Metzinger. «Ser nadie» no puede ser más que una posición epistemológica de cara a la investigación filosófica y científica de la mente. Al contrario que otros autores materialistas, Metzinger reconoce la autorrefutabilidad de su tesis última. Solo cobra sentido el concepto de «ser nadie» en relación con un proyecto de investigación científica y filosófica. Que Metzinger, según Murillo, rechazara una teoría verificacionista de la verdad en favor de una coherentista y pragmatista no es solo una cuestión secundaria, sino

que es precisamente lo que otorga sentido a la proposición de «no somos nadie», pues tiene consecuencias en la práctica. Esto, evidentemente, también tiene consecuencias éticas, que desarrollaremos posteriormente.

El principal fallo de la crítica de Murillo es que no pareció comprender que Metzinger rechaza el reduccionismo epistemológico, aunque no el reduccionismo ontológico, que podría haber sido el objeto de su crítica. Precisamente, Murillo hace una crítica que podría aplicarse a la filosofía de Daniel Dennett, pero no a la de Metzinger.

La diferencia principal entre la filosofía de Metzinger y la de Dennett es que para Dennett los datos primarios que tiene que explicar una teoría de la consciencia son «los datos primarios son los sonidos que se registran cuando los sujetos mueven la boca, o los diversos enunciados (...) sin interpretar» (Dennett, 2006: 61). La neurofenomenología de Metzinger se toma en serio los datos fenoménicos y busca explicar ese dato consciente cuyo conocimiento es irreductible. Sin embargo, para superar el problema de la inefabilidad y el lenguaje privado, define estos estados a partir de estados cerebrales que se pueden observar públicamente. Pese a que se inspira en la metodología de Dennett, no saca las mismas conclusiones y busca «tomarse lo fenoménico en serio» (Weisberg, 2005: 8). La forma de tomar los datos de Dennett también es semejante a la usada por Jackson y Chalmers, quienes proponen una ontología completamente distinta tanto a Dennett como a Metzinger (Weisberg, 2005: 8).

El materialismo de Metzinger no encaja directamente en el materialismo de «tipo A» o «tipo B» según la clasificación de Chalmers (Chalmers, 2003; Tomasik, 2018). Frente a los del tipo A, Metzinger rechaza la reductibilidad conceptual y epistémica de las experiencias conscientes y se opone al funcionalismo o conductismo analíticos. También considera como inteligible la pregunta de si determinado ser es consciente, es decir, si posee un automodelo fenoménico. Precisamente, lo que busca es definir neurofenomenológicamente el automodelo para poder responder esta pregunta de cara a seres artificiales. Metzinger, frente a los materialistas del tipo B, rechaza la inteligibilidad de los «zombis» y la existencia de la «brecha explicativa».

Precisamente, la filosofía de Metzinger permite responder a las críticas habituales al materialismo, como la que hace Murillo. Metzinger define los «qualia» a través de estados neuronales públicamente observables,

tomándose en serio la subjetividad, pero rechazando la infabilidad. La dificultad reside en compaginar efectivamente este «tomarse en serio la subjetividad» sin rehabilitar la brecha explicativa. La propuesta de Metzinger, precisamente, permite un posible acercamiento que logra esto.

Otra crítica es la de Graham Harman, que lleva la atención a que Metzinger no parece llevar hasta sus últimas conclusiones su argumento. En vez de titular a su libro «Ser nadie», podría haberlo titulado «Ser nada». Si hemos eliminado la realidad ontológica del «yo», que es un mero automodelo, un sistema físico que nos presenta una realidad ilusoria; esa realidad que se nos presenta es también ilusoria. Es decir, no solo es ilusorio el «yo», que es simplemente el sistema mediante el cual se representa; cualquier objeto fenoménico pierde su realidad ontológica de la misma manera que el «yo» (Harman, 2011: 25-26).

Metzinger aceptaría esta crítica y afirmaríala sin problemas junto a Sellars que:

el mundo perceptivo es fenoménico en un sentido análogo al kantiano; y (...) el mundo real o “nouménico”, sobre el que se apoya el “mundo de las apariencias”, no es un mundo metafísico de cosas incognoscibles en sí mismas, sino simplemente el que construye como interpretación la teoría científica (Sellars, 1971: 108).

A pesar de que no se negaría la experiencia subjetiva y se trataría como propiedades de determinados sistemas que constituyen automodelos.

Otra crítica que hace Harman es que la conclusión de que no hay «yo» ya está implícita en su materialismo reductivo. Para Harman, aunque el yo estuviera causado enteramente por procesos «físico-orgánicos-como-un-todo», desde su ontología orientada a objetos, eso no implicaría que lo único realmente existente fueran esos procesos físico-orgánicos, sino también el «yo». El hecho de que algo esté formado por otros «objetos» no implica que eso no sea un objeto ontológicamente real. Para Graham Harman:

[Esta] visión devastadora de un mundo «nemocéntrico» aterrador, donde los yoes no existen, [se basa] en el simple dogma *a priori* de que, si algo tiene antecedentes causales, entonces solo aquellos antecedentes pueden tener una realidad independiente. Pero esto es simplemente una familiar y trivial forma de materialismo aplicado a la neurofísica, no una visión filosófica de ningún tipo de rigor pionero (Harman, 2011: 28-29).

El problema de esta crítica es que Metzinger no está solo descubriendo los antecedentes causales del «yo», sino que está disolviendo su existencia ontológica independiente y reduciéndola a procesos físicos con determinadas propiedades fenoménicas. No es un «dogma» *a priori* sino la disolución de un concepto cargado y la defensa de una ontología materialista que permite nuevas propuestas para la investigación filosófica y científica (Metzinger, 2003: 627-628).

Hemos visto en esta sección como las críticas a la filosofía de Metzinger no tienen éxito. Además, muchas de ellas son críticas dirigidas al materialismo en general, que se pueden aplicar a muchas teorías filosóficas materialistas, pero que Metzinger contesta explícitamente. La propuesta de Metzinger tiene la dificultad de encajar «tomarse en serio la subjetividad» con no rehabilitar la brecha explicativa, sin embargo, ninguna de las críticas se ha dirigido a este punto. Por ello, la propuesta de Metzinger se muestra como una opción viable.

5. SOBRE EL LIBRE ALBEDRÍO

El tratamiento de la libertad que hace Metzinger es principalmente empírico, es decir, estudia las condiciones de posibilidad para que surja la experiencia de la voluntad y el libre albedrío, de este estudio deducirá las consecuencias filosóficas. Metzinger recuerda que en su teoría del automodelo la experiencia más simple del «túnel del yo» no incluye necesariamente la agencia. Esta es, por ejemplo, la diferencia entre un sueño y un sueño lúcido, en ambos hay un túnel del yo, pero solo en el último se experimenta la agencia.

Otro ejemplo que Metzinger considera importante es el de la «mente errante», nuestra mente divaga durante aproximadamente el 50% del tiempo de vigilia. Estos son momentos en los cuales perdemos nuestra autonomía mental, por ejemplo, cuando nuestros ojos pasan por las líneas de un libro sin captar lo que dice. Fenoménicamente, no se puede captar el primer pensamiento que nos lleva hasta esta ensoñación o divagación interna, momento en el cual colapsa el «modelo del agente epistémico». Es decir, se ha descompuesto una parte de nuestro modelo, la agencia (Metzinger, 2018: 167).

Frente al concepto de «sujeto cognitivo» tradicional de la filosofía, Metzinger habla de la «agencia cognitiva». Durante la vigilia es muy habitual la mente errante, los pensamientos pasan por nuestro automodelo sin experimentar la agencia. La visión tradicional del «sujeto cognitivo», es decir, un «yo» que piensa esos pensamientos solo ocurre cuando en nuestro automodelo se genera una «agencia cognitiva». Metzinger lo compara con el concepto de Husserl de «relámpago de la atención» (*Blickstrahl der Aufmerksamkeit*), es decir, esta agencia vendría por la experiencia fenoménica de ser capaz de controlar este «relámpago de la atención» (Metzinger, 2018: 168).

Una vez se ha conceptualizado la experiencia de la agencia, Metzinger nos pone de ejemplo un experimento en el cual se estimuló el banco ventral del surco cingulado anterior del cerebro a una paciente. Esta paciente, fenoménicamente, sintió la necesidad de agarrar un objeto, movió su mano y lo agarró. Metzinger concluye a partir de esto: «sea lo que sea la experiencia consciente del libre albedrío, parece algo que puede conectarse y desconectarse con la ayuda de la pequeña corriente eléctrica de un electrodo en el cerebro». El autor pone ejemplos de otros experimentos en los cuales se «alucina» la agencia, ya sea a través de electrodos, sugestión, etc. La agencia surge, simplemente, de vincular fenoménicamente la captación interna de la intención con la representación de los movimientos presentes (Metzinger, 2018: 175-179).

El tratamiento que le da al libre albedrío es esperable: «El estado presente del universo físico siempre determina al siguiente, y el cerebro es parte de este universo». Metzinger afirma el determinismo y dice que «se trata, sencillamente, de la visión científica del mundo». También rechazan todas las teorías cuánticas acerca del libre albedrío por dos motivos: que algo sea completamente azaroso deja tan poca libertad al ser humano como la determinación total; además, simplemente, los efectos cuánticos no se dan en objetos tan grandes y calientes como una célula nerviosa a 37 grados. Metzinger no niega la experiencia fenoménica de que en cualquier momento podamos iniciar una cadena causal y la fuerte sensación de que, en las mismas circunstancias, podríamos actuar de otra manera. Sin embargo, para Metzinger esto no se corresponde con lo que sabemos del cerebro y del mundo físico; además, se acaba de explicar neurofenomenológicamente de dónde surge esa sensación de agencia (Metzinger, 2018: 181-184).

Metzinger lleva también la atención al hecho de que la fenomenología de la voluntad no es solo algo individual y personal; sino que está impartida socialmente con todos los conceptos de castigo, responsabilidad, etc., que refuerzan la *apariencia* de la libertad, aun cuando fenomenológicamente es más escurridiza de lo que parece (Metzinger nos insta a prestar atención a la experiencia fenoménica de mover un brazo intencionadamente para hacer ver cómo de escurridiza es la causalidad y cómo es imposible de captar fenoménicamente de forma clara el concepto dualista de «una acción mental causa un evento físico»). Metzinger predice que cuando avance la neurociencia y se diluciden los mecanismos que generan nuestros deseos, intenciones y acciones, la noción de libre albedrío perderá peso. Hasta el punto de que «el castigo aparecerá como un concepto primitivo, heredado de un pasado animal» y «[veremos] la experiencia subjetiva del libre albedrío como una ingenua herramienta neurocomputacional» (Metzinger, 2018: 182-185).

Antes mencionamos que este era el tratamiento esperable del libre albedrío, sin embargo, esto no es del todo cierto, porque desde otras perspectivas materialistas se le da en ocasiones otro tratamiento distinto: la redefinición o la emergencia. La redefinición es lo que hacen los compatibilistas, quienes afirman que el determinismo y el libre albedrío son compatibles. Por ejemplo, Thomas Hobbes, materialista mecanicista que afirmaba el determinismo completamente, definía «libertad» como ausencia de oposición; de este modo, un hombre es tan libre y está tan determinado como un río en el que no hay presas. Otra de las personas que defiende esta visión es Daniel Dennett, que considera peligroso afirmar que el libre albedrío no existe, porque tiene consecuencias en la forma de actuar de las personas, ya que va en contra de nuestras nociones habituales de responsabilidad. El libre albedrío queda definido como una simple «autonomía», la acción es controlada por «nosotros mismos» y no actuamos «manipulados por otros» (Dennett, 2016: 125-133).

El primer problema de esta visión es que, aun aceptando la premisa, la cadena causal acabaría por salir de nosotros, aunque, en algún sentido seamos causa de nuestras acciones, jamás seríamos causa de cuándo y dónde nacimos, nos criamos, de quiénes somos hijos, etc. Esta visión no sería, ni siquiera, coherente internamente. Además, la noción de libre albedrío ha quedado definida hasta el punto de que es completamente trivial,

irrelevante y no tiene nada que ver con la noción de libre albedrío que nos dota de responsabilidad y agencia moral. La existencia del libre albedrío de Dennett es completamente irrelevante. En vez de hacer como Metzinger, ser coherente con su ontología, negar el libre albedrío y afirmar que, si de verdad no existe, se deberían replantear completamente los conceptos de responsabilidad, agencia y moral y castigo. Dennett afirma algo semejante a: el libre albedrío verdadero es falso, pero esto no tiene consecuencias de ningún tipo, porque el libre albedrío falso es verdadero.

Edelman defiende que gracias a la complejidad del sistema nervioso surge una autoconciencia de carácter simbólico de la cual emerge la «libertad». Damasio apela a la capacidad del ser humano para enfrentarse a su biología y su cultura para hablarnos de «libertad». El problema de esta articulación es la siguiente: si no se está redefiniendo la «libertad» como hacen los compatibilistas, se está cayendo en un error categorial. Este error categorial viene del hecho de creer que de la complejidad de un sistema físico pueda emerger la libertad. El debate del libre albedrío se resuelve instantáneamente cuando se opta por una ontología referente al problema mente-cerebro. Los procesos físicos están determinados causalmente, de ninguna manera son libres; por muy imprevisible y complejo que sea un sistema físico, en absoluto es libre. La pregunta sobre el libre albedrío no es física (ni biológica, ni neurológica, etc.), es ontológica: una vez hemos caracterizado la ontología de la mente, la pregunta está respondida.

El tratamiento de la libertad de Metzinger es interesante debido a su claridad y rotundidad; una vez se ha afirmado que la consciencia es un proceso físico, se tiene que negar el libre albedrío. Frente a Dennett, que trata de rescatar el libre albedrío redefiniéndolo, Metzinger llega a proclamar un futuro cambio social en torno a la inexistencia del libre albedrío.

6. LA PROPUESTA ÉTICA

La ética de Metzinger se puede caracterizar como un utilitarismo *sui generis* que considera moralmente a todos los túneles del «yo» (humanos, animales y artificiales). Para empezar, vamos a ver cómo discute Metzinger la posible existencia de *robots* conscientes. Desde el monismo materialista de Metzinger, la respuesta es esperable:

Si un sistema puede integrar una imagen interna transparente de sí mismo con esta realidad fenoménica, entonces se aparecería a sí mismo. Se convertirá entonces en un yo y un realista ingenuo sobre lo que su automodelo le diga que hay. (...) Creará su propia existencia (Metzinger, 2018: 260).

Al no haber nada inmaterial en la consciencia, ciertos procesos físicos conllevan, necesariamente, la existencia de una «primera persona». Lo que diferencia la visión de Metzinger de otras visiones sobre la inteligencia artificial es que no es en absoluto cognitivista. Frente a los modelos donde se trata de imitar los procesos humanos, que las máquinas pasen el test de Turing, etc., para Metzinger lo que hace que algo sea consciente es su capacidad para generar un automodelo fenoménico; no tiene nada que ver con la inteligencia o las capacidades cognitivas. El proyecto de la máquina consciente es perfectamente posible ontológicamente, la dificultad es técnica y su solución la otorgaría la investigación de la consciencia. Una vez se ha hecho una definición precisa de en lo que consiste la subjetividad, se podría incorporar en una máquina (Metzinger, 2018: 262).

La ventaja de la visión de Metzinger es que lo libera de muchas objeciones. Por ejemplo, la objeción de la habitación china se dirige a un criterio basado en la inteligencia, mientras que la existencia del automodelo fenoménico de Metzinger depende de procesos desligados de la inteligencia.

Metzinger da un giro hacia la ética. ¿Deberíamos crear estas máquinas artificiales del yo? Metzinger responde con un rotundo no; estas máquinas, como sujetos con experiencia subjetiva, capaces de sufrir, serían objetos de consideración moral. Es posible que se pudiera crear estas máquinas antes de conocer con seguridad si estas de verdad sufren, cómo sería la experiencia del sufrimiento; correríamos el riesgo de crear unas máquinas del «yo» con una existencia miserable (Metzinger, 2018: 263-265).

El neurofilósofo continúa con este experimento mental, ¿y si fueran Máquinas Dichosas artificiales? Es decir, algo similar a la Máquina de Experiencia de Nozick, solo que la máquina sería el propio sujeto de la experiencia, no un humano enchufado a ella. Metzinger responde que tampoco deberíamos crear estas máquinas. Primero, estas Máquinas Dichosas podrían ser defectuosas, que la aparente experiencia fenoménica deseable en realidad sea miserable pero que no tengamos forma de saberlo. Además, probablemente replicaríamos la estructura psicológica humana, la

cual está caracterizada por el dolor y el aburrimiento. Metzinger acaba por defender una postura que llama «pesimismo fenomenológico»:

La variedad de experiencia fenoménica generada por el cerebro no es un tesoro sino una carga: sopesados en la medida de una vida, la balanza entre alegría y sufrimiento se inclina hacia el último en casi todos los casos. Desde Buda a Schopenhauer, existe una larga tradición filosófica que afirma, fundamentalmente, que la vida no merece la pena ser vivida (Metzinger, 2018: 268).

Metzinger continúa articulando este pesimismo y afirma que hay un punto ciego en la fenomenología y la filosofía de la mente, «el sufrimiento»:

Se han escrito miles de páginas sobre *qualia* de color y sobre *zombies*, pero prácticamente ninguno se ha dedicado a estados fenoménicos ubicuos como el dolor físico, el aburrimiento o la (...) depresión subclínica (Metzinger, 2018: 268).

Metzinger se está inscribiendo en una larga tradición de pesimismo. Afirma que la evolución no persigue fines, simplemente ocurre, y ha dado lugar a una estructura fenoménica caracterizada por el sufrimiento. Este planteamiento recuerda enormemente a la voluntad de Schopenhauer, una voluntad ciega que simplemente quiere la vida y que genera seres caracterizados por deseos insatisfechos, dolor y aburrimiento. Pese a que Metzinger sustituye la explicación metafísica por una naturalista y darwinista, la conclusión es la misma: la naturaleza ha generado unos seres con unas experiencias fenoménicas caracterizadas por la ubicuidad del dolor y el aburrimiento. La selección natural no seleccionó a los individuos más felices, sino a los que fueron capaces de sobrevivir; la evolución ha seleccionado a unos individuos enfrascados en una rueda del hedonismo.

Metzinger solo considera moralmente a los estados fenoménicos. El ser humano tiene consideración moral porque es una «máquina de *yo*», al igual que los animales y los supuestos *robots* conscientes. Metzinger propone el «principio ético de minimización del sufrimiento» el cual «establece que debemos reducir los estados fenoménicos negativos en todos los seres conscientes con capacidad de sufrir». A su vez, los estados fenoménicos negativos se definen como «estados de conciencia desagradables o angustiosos» (Metzinger, 2018: 331-332).

El planteamiento ético es un utilitarismo en el que se le da más valor a la reducción de sufrimiento que a la creación de placer (a la cual también le da valor) y que considera moralmente a todas las experiencias

fenoménicas. Metzinger, al haber caracterizado de esta forma pesimista la experiencia fenoménica, no le da prioridad a la búsqueda de una experiencia fenoménica positiva, sino a evitar una experiencia fenoménica negativa. Metzinger considera éticamente también a los seres conscientes futuros y ha apoyado previamente evitar la creación de máquinas del «yo» artificiales, es decir, considera inmoral traer seres sintientes artificiales al mundo (Metzinger, 2018: 325-332). Por ello, Metzinger propone un moratorio contra la investigación que puede provocar la emergencia de seres conscientes artificiales (Metzinger, 2021).

Metzinger introduce el concepto de una superinteligencia benevolente que conoce todas las experiencias fenoménicas y aplica la máxima utilitarista de minimizar el sufrimiento y aumentar la felicidad. Esta máquina concluye que la existencia está mucho más llena de sufrimiento y de insatisfacción que de placer y, aplicando el principio de utilidad, concluye que los seres sintientes deberían dejar este mundo de forma pacífica (Metzinger, 2017). Para Metzinger, esto es más bien un ejercicio acerca de la inteligencia artificial que un experimento ético, pero en este experimento se muestra favorable a la posición antinatalista, en contra de todos los sesgos humanos que surgieron de la evolución y nos hacen perpetuarnos a nosotros mismos y a nuestra especie en la existencia, en contra de nuestros propios intereses.

Además de esto, Metzinger, a raíz de diversos estudios neurológicos, sugiere la meditación y el autoconocimiento como medidas para tener mayor autonomía mental, aumentar los estados fenoménicamente positivos, reducir los momentos en los que nuestra mente divaga y perdemos la experiencia fenoménica de nuestra agencia (Metzinger, 2018: 337-341). También introduce un criterio ético similar al eterno retorno de Nietzsche; una «playlist eterna» donde podemos elegir las experiencias fenoménicas que viviremos tras la muerte. Metzinger realizó un experimento donde se llamaba aleatoriamente a los sujetos para preguntar si incluirían la experiencia fenoménica que estaban viviendo en ese momento en esa «playlist» y solo una media de 11-13% de las muestras fenoménicas se incluyeron en esta línea. El pesimismo fenomenológico de Metzinger aparece apoyado empíricamente (Metzinger, 2018: 325-326).

Metzinger se preocupa de vincular su filosofía de la mente con consideraciones prácticas. Los principales puntos son este pesimismo fenomenológico y la defensa del merecimiento de consideración

moral de las posibles inteligencias artificiales conscientes. También podríamos notar que la disolución del concepto del «yo», apoyada en la neurofenomenología, tiene resonancias con la discusión presente en *Razones y personas* de Derek Parfit, que critica la teoría del interés egoísta y defiende la racionalidad del altruismo basándose en una visión reduccionista de la identidad¹. Además, también tiene claras consecuencias éticas su ataque al libre albedrío. La preocupación ética de Metzinger es clara.

7. CONCLUSIONES

En el presente artículo hemos expuesto y defendido la filosofía de la mente de Thomas Metzinger. Primero hemos examinado el método neurofenomenológico, que define estados mentales en función de correlatos neuronales. Metzinger busca tomarse en serio la subjetividad e investigar los correlatos neuronales que la permiten. A su vez, critica el materialismo eliminativo por disolver la experiencia subjetiva y considera que la experiencia fenoménica nos da un conocimiento epistémico irreductible al conocimiento de procesos físicos. Sin embargo, los *qualia* definidos como privados e inefables son incoherentes, por ello, los define en función de correlatos neuronales.

A partir del estudio de determinadas experiencias (sueños lúcidos, experiencias extracorporales, etc.), propone que el rasgo necesario y suficiente de «ser un yo» es que se genere un automodelo fenoménico, que permite la «aparición de un mundo interno». Este automodelo es transparente, es decir, no podemos saber en primera persona cómo se genera. Aquí surge el concepto de «túnel del yo», un túnel de experiencias fenoménicas que recorreremos sin ver sus paredes. Sin embargo, en sentido estricto, *nadie* recorre el túnel, sino que esas experiencias fenoménicas llevan consigo la existencia de la experiencia de «auto-identificación» y «auto-localización». Este automodelo nos hace «realistas ingenuos», pues se genera un mundo aparente habitado por un yo en contacto directo con determinados aspectos de la realidad.

Tras esto, examinamos unas críticas insuficientes a la filosofía de Metzinger. Un error habitual en las críticas hacia Metzinger es que son

1. Se desarrolla esta comparación en (Maroufkhani, 2017: 93 y ss).

críticas genéricas a proyectos materialistas (funcionalistas, eliminativistas, etc.), no directas a la obra de Metzinger, por ello, muchas de estas críticas ya están contempladas en la propia obra. Una dificultad del proyecto de este autor es combinar «tomarse en serio la experiencia fenoménica» con no rehabilitar la brecha explicativa. El autor rechaza la existencia de esta brecha en un sistema, pero es un punto que queda por tratar exhaustivamente.

Finalmente, hemos examinado las consecuencias éticas de su filosofía. Metzinger afirma tajantemente la inexistencia del libre albedrío y, frente a los compatibilistas que redefinen la agencia, busca que la negación del libre albedrío tenga consecuencias prácticas en el tratamiento interpersonal. Metzinger es un pesimista fenomenológico que afirma que, en general, las experiencias conscientes son en su mayoría negativas, por ello, se opone a la creación de máquinas conscientes.

REFERENCIAS BIBLIOGRÁFICAS:

- Arana, J. (2015): *La conciencia inexplicada*. Madrid: Biblioteca Nueva.
- Blanke, O., & Metzinger, T. (2009): «Full-body illusions and minimal phenomenal selfhood». *Trends in cognitive sciences*, 13(1), 7-13.
- Chalmers, D. J. (2003): «Consciousness and its place in nature». *Blackwell guide to the philosophy of mind*, 102-142.
- Dennett, D. (2006): *Dulces sueños: Obstáculos filosóficos para una ciencia de la conciencia* (J. Barba & S. Jawerbaum, Trads.). Buenos Aires: Katz Discusiones.
- Dennett, D. (2016): «Free Will Worth Having». En D. Edmonds & E. Warburton, *Philosophy Bites Again* (Oxford University Press: 125-133).
- Harman, G. (2011): «The Problem With Metzinger». *Cosmos and History: The Journal of Natural and Social Philosophy*, 7(1), 7-36.
- Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007): «Video ergo sum: Manipulating bodily self-consciousness». *Science*, 317(5841), 1096-1099.
- Maroufkhani, K. P. (2017): *Selfhood and the Metaphysics of Altruism*. Universidad de Hawái, Manoa.
- Metzinger, T. (2003): *Being No One: The Self-Model Theory of Subjectivity*. The MIT Press. <https://doi.org/10.7551/mitpress/1551.001.0001>
- Metzinger, T. (2017): «Benevolent Artificial Anti-Natalism». *Edge*.

Recuperado de https://www.edge.org/conversation/thomas_metzinger-benevolent-artificial-anti-natalism-baan

Metzinger, T. (2018): *El túnel del yo: Ciencia de la mente y mito del sujeto*. Madrid: Enclave de libros.

Metzinger, T. (2021): «Artificial suffering: An argument for a global moratorium on synthetic phenomenology». *Journal of Artificial Intelligence and Consciousness*, 8(01), 43-66.

Murillo, J. I. (2011): «Apariencia y realidad del yo: Una aproximación crítica a la propuesta de Thomas Metzinger». En *El asalto a lo mental: Neurociencias, consciencia y libertad* (pp. 113-124). Madrid: Biblioteca Nueva.

Sellars, W. (1971): *Ciencia, percepción y realidad*. Madrid: Tecnos.

Tomasik, B. (2018): «Is There A Hard Problem of Consciousness?». *Essays on reducing suffering*. Revisado el 1 de junio de 2022, <https://reducing-suffering.org/hard-problem-consciousness/>

Weisberg, J. (2005): «Consciousness constrained». *Psyche*, 11(5), 1-20.

JOSÉ CURBUERA-LUIS es estudiante de Filosofía en la Universidad de Salamanca.

Líneas de investigación:

Ética animal, éticas centradas en el sufrimiento, consecuencialismo.

Publicaciones recientes:

- Curbera Luis, J. «Un Enfoque Principialista Respecto a Los Xenotrasplantes», *Dilemata*, n.º 42, septiembre de 2023, pp. 13-29.

Correo: joseramoncurbera@usal.es

