Neuromodulación para la mejora de la agencia moral: el neurofeedback*

Neuromodulation for the Enhancement of Moral Agency: Neurofeedback

Paloma García Díaz
Universidad de Granada
palomagdiaz@ugr.es

ISSN 1989-7022

Resumen: Este artículo defiende que el proyecto de mejora moral precisa de una mayor atención a las dimensiones racionales y deliberativas de la agencia moral. Para ello, se presenta la contribución del neurofeedback a la mejora de dichas deliberaciones morales y de la autonomía. Esta interfaz cerebro-ordenador, por lo demás, se toma como un posible componente de un asistente moral socrático (Lara y Deckers 2020) que vela porque la mejora moral se produzca gracias a una interacción plena entre los agentes morales y dicho asistente. Esta propuesta se aleja del proyecto de biomejora moral a través de la neurofarmacología, centrado en la mejora de las emociones. Asimismo, se distancia de la idea de delegar la toma de decisiones morales en las agencias morales artificiales, lo que supondría apostar por un modelo continuista entre las agencias morales humanas y artificiales.

Palabras clave: Mejora moral, agencia moral, neurofeedback, modelo de interacción plena, agencias morales artificiales

Abstract: This article aims to pay heed to the rational and deliberative dimensions of moral agency within the project of moral enhancement. In this sense, it is presented how the technique of neurofeedback might contribute to the enhancement of moral deliberations and autonomy. Furthermore, this brain-computer interface is thought as a possible element of a Socratic moral assistant (Lara & Deckers 2020) interested in improving moral enhancement within a model of full interaction between moral agents and such a moral assistant. This proposal does not embrace the project of moral bioenhancement, centred in the improvement of moral emotions through neuropharmacology. Moreover, it rejects the delegation of moral decision-making to moral artificial agents because this would mean that there is continuity between moral human agents and artificial moral agents.

Keywords: Moral enhancement, moral agency, neurofeedback, full interaction model, artificial moral agents

Paloma García Díaz:
"Neuromodulación para la mejora de la agencia moral: el neurofeedback", en Jon Rueda (ed.): *Tecnologías socialmente disruptivas*
ILEMATA, Revista Internacional de Éticas Aplicadas, nº 34, 105-119

1. Revisando el proyecto de la mejora moral: los enfoques de bio e info-mejora

La necesidad de hacer frente a los grandes problemas de las sociedades –como la pobreza, el cambio climático, la defensa de la democracia o las pandemias– está a la base del interés en ética por el proyecto de la mejora moral. Desde este, en síntesis, se plantea cómo las personas podrían mejorar moralmente para poder hacer frente a los grandes retos a los que nos enfrenta la realidad actual. El proyecto de mejora moral trata, pues, de que sean superadas las limitaciones de la psicología humana que nos impiden hacer frente a los desafíos globales. (Persson y Savulescu 2014, Savulescu y Maslen 2015)[1].

El creciente interés por el tema de la mejora moral en el terreno académico, por lo demás, se ha abordado a través de biomejoras e infomejoras. La biomejora moral se han planteado desde técnicas mundanas y disponibles, como la neurofarmacología, o técnicas más sofisticadas como la ingeniería genética. En el terreno de la biomejora se ha explorado cómo mejorar las emociones, o cómo hacer decrecer las emociones morales negativas –como el exceso de agresividad– mediante técnicas farmacológicas (Persson y Savulescu 2014; Douglas 2008 y 2013). Douglas (2008, 229), en este sentido, considera que la mejora moral consiste en alteraciones para tener mejores motivos morales.

* Este trabajo forma parte del Proyecto de Investigación de "Éticas digitales. Mejora Humana desde el uso interactivo de la Inteligencia Artificial" (ID2019-104943RB-I00 / SRA (State Research Agency / 10.13039/501100011033)

Agradezco a las revisoras y/o revisores de este artículo por sus comentarios y sugerencias de mejora

Received: 08/12/2020
Accepted: 12/12/2020

ISSN 1989-7022

Dilemata, año 13 (2021), nº 34, 105-119

Este autor insiste en que la mejora de las disposiciones morales debe entenderse como una disminución de emociones negativas, como la aversión a grupos raciales o los impulsos hacia agresiones violentas (Douglas 2008, 231). En efecto, estudios sobre técnicas neurofarmacológicas muestran algunos resultados positivos en el ámbito de la mejora moral, como por ejemplo las intervenciones farmacológicas con oxitocina (Dubljević y Racine 2017, 344). A esta sustancia se la conoce como la "molécula moral", pues las investigaciones muestran que la oxitocina favorece la conducta pro-social, aumentando la cooperación y la confianza entre los miembros del grupo. Aunque la oxitocina, como efecto colateral, no consigue que los sujetos se vuelvan más empáticos y compasivos con miembros de otros grupos.

Así pues, este proyecto de biomejora no ha estado exento de críticas. Por citar algunas de ellas –además de los efectos colaterales de las técnicas– cabe mencionar, primero, que se evalúa con escepticismo la idea de mejorar moralmente a los individuos como estrategia para hacer frente a problemas estructurales de carácter nacional, transnacional y global como el problema de la pobreza mundial (De Melo-Martín y Salles 2015, 227; Harris 2013b, 290). Segundo, se ha argumentado en contra de abordar la mejora moral exclusivamente en términos de mejora de las emociones como, por ejemplo, la empatía o la compasión. Esta circunstancia por la que se obvian los aspectos racionales de la moral podría derivar en un socavamiento de la libertad y la autonomía y, además, podría ser nocivo para la agencia moral (Harris 2013a y 2013b; DeGrazia 2013).

Además de la perspectiva neurofarmacológica, el proyecto de la mejora moral ha experimentado un gran desarrollo en el terreno de las infotecnologías. En concreto, esto se ha producido en el campo de las máquinas éticas o agentes morales artificiales (AMAs en adelante) (Moor 2006; Wallach 2010; Cervantes *et al.* 2020). Bajo la denominación de AMAs se agrupan diferentes tecnologías digitales y robots que incorporan en su diseño, y despliegan en su funcionamiento, principios éticos. Algunos de estos principios sobrepasan meramente el no actuar inmoralmente y, por tanto, estos agentes incorporan mecanismos de toma de decisiones desde principios morales. Algunos AMAs son defendidos, pues, como mentores o guías respecto de qué decisiones éticas serían más convenientes adoptar en ciertas circunstancias, tomando como referencia ciertos principios morales (Fossa 2018). Diferentes autores sostienen, a este respecto, que los agentes morales artificiales nos permiten comprender mejor qué es la ética y la moralidad humana (Coekelbergh 2010, 240; Wallach 2010, 244). Y, desde este enfoque, se defiende que hay una continuidad entre los agentes morales y los agentes morales artificiales, es decir, que no hay ninguna diferencia cualitativa entre ambas entidades. Además, se considera que los agentes morales humanos y los agentes morales artificiales deberían ser concebidos como entidades homogéneas (Fossa 2018, 115). Así pues, desde este enfoque continuista se aborda la mejora moral como una posibilidad que propician dichos AMAs, los cuales, en algún estadio de su desarrollo, sobrepasarán incluso las limitaciones de las deliberaciones morales de los seres sintientes (Wallach 2010, 249). Estos AMAs perfilan, pues, un nuevo escenario para la mejora moral y, por ende, para la ética en general. Sus defensores, además, confían de manera optimista en que los avances en el diseño y funcionamiento de estas máquinas contribuirá a una nueva y mejor definición de qué es la ética y a una superación de las deliberaciones éticas humanas. Algunos ejemplos de AMA son los siguientes: *GenEth* es un analizador general de dilemas éticos basados en conocimientos sobre características éticamente relevantes, obligaciones, acciones, casos y principios. *GenEth* opera con una programación de lógica inductiva. Por otro lado, *Jeremy* y *W.D.* son dos

ISSN 1989-7022

Dilemata, año 13 (2021), nº 34, 105-119

máquinas de aprendizaje automático (*machine learning)* para resolver dilemas éticos y están basadas en principios éticos normativos procedentes del utilitarismo (Cervantes *et al.* 2020).

En este artículo, tras haber presentado brevemente el proyecto de mejora moral desde técnicas neurofarmacológicas, se discute con la propuesta de la delegación de la toma de decisiones éticas en robots o inteligencias artificiales éticas. Para ello se van a analizar los problemas que se derivarían para la agencia moral humana si se aceptase el enfoque continuista. Por último, se presenta la apuesta por una mejora tecnológicamente mediada de las deliberaciones morales gracias a la técnica no invasiva del *neurofeedback*. Esta es una técnica que requiere un aprendizaje activo por parte de los sujetos de sus propias ondas cerebrales. Esta práctica de neuromodulación trae consigo mejoras en el ámbito del control emocional, la concentración, la atención, la memoria, la creatividad y propicia mejores desempeños cognitivos y motores (Enriquez-Geppert *et al.* 2017b).

El *neurofeedback*, asimismo, se considerará como un posible componente de un proyecto más complejo de mejora moral humana, a saber, una inteligencia artificial ética, denominada "asistente socrático para la mejora moral" (Lara y Deckers 2020). Este proyecto apuesta por una mejora de los aspectos racionales de la deliberación moral. Consiste, brevemente, en plantear una interacción continuada entre los agentes morales y un asistente gracias a la cual el sujeto vaya contrastando sus ideas, creencias y valores con información empírica relevante. Asimismo, la interacción con el asistente trabajaría la claridad conceptual y perfilaría las argumentaciones morales. Este asistente no sucumbe a la idea sostenida, desde la tesis continuista, según la cual los agentes humanos deberían delegar toda su capacidad de deliberación moral a una inteligencia artificial ética. Esto constituiría, según Lara y Deckers (2020), un proyecto exhaustivo de mejora moral por el que, lejos de potenciar las deliberaciones y la autonomía de los agentes morales, se minaría dicha autonomía.

En las líneas que siguen, pues, se presentan algunos rasgos de los AMAs y se problematiza la idea de que en un futuro serán superiores, sin más, a los agentes morales humanos[2]. Tras esto se realizan algunas reflexiones respecto de la naturaleza de las agencias morales humanas y artificiales para, posteriormente, presentar el posible refuerzo de la agencia moral humana gracias a la interfaz cerebro-ordenador del *neurofeedback*.

2. El enfoque continuista en la mejora moral

2.1. ¿Pueden considerarse como agentes morales solo a los seres humanos?

La creciente mediación digital y automatización de nuestras prácticas con la inteligencia artificial, bien sea mediante *softwares* o robots, en el terreno de la medicina, el cuidado, los negocios o la guerra suscita múltiples cuestiones relativas a los aspectos éticos de las diferentes infotecnologías y máquinas (Bryson 2010, 2018; Brey 2018; Cervantes *et al.* 2020; Coeckelbergh 2018; Gunkel 2020). Estos interrogantes son especialmente interesantes cuando se abordan las diferentes máquinas éticas, robots éticos o AMAs.

La aparición de los AMAs en escenarios dispares se presenta como una nueva vía de mejora de las deliberaciones morales y de la ética en general. Desde el proyecto de mejora moral gracias

ISSN 1989-7022

Dilemata, año 13 (2021), nº 34, 105-119

a las infotecnologías, se pone el foco en las limitaciones de los agentes morales humanos a la hora de deliberar y actuar moralmente. Así pues, se considera que los diseños computacionales que incorporan los AMAs superan los déficits de los seres humanos a la hora de razonar y tomar decisiones éticas. La moralidad se concibe, desde esta perspectiva, desprovista de afectividad y emociones. La mejora de la empatía y de la compasión, requisitos destacados en buena parte de los defensores de la biomejora moral, no son contemplados como requisitos de la ética. Solo desde esta perspectiva que elimina la afectividad se puede hablar de una posible continuidad entre los agentes morales humanos y artificiales. Con la aparición de los AMAs, pues, los agentes morales parecen haber proliferado y haberse hecho más complejos. Frente a las limitaciones mencionadas previamente de los proyectos de la biomejora moral, un nuevo aliado de la mejora moral parece haberse encontrado en las promesas de los nuevos AMAs.

Una interesante revisión respecto del estado de la cuestión en el desarrollo y clasificación de los diferentes AMAs se encuentra en el trabajo de Cervantes y colaboradores (2020). Para estos autores, un agente moral artificial puede ser tanto un agente virtual (*software)* como un agente físico (robot). Se caracteriza por actuar moralmente o por evitar una conducta inmoral. Las conductas de estos agentes morales pueden estar, aunque no necesariamente, basadas en los principios de algunas de las teorías éticas más relevantes, como las teorías éticas teleológicas, deontológicas y la ética de la virtud, (Cervantes *et al*. 2020, 505).

La detallada revisión de los AMAs de Cervantes *et al*. (2020) pone el acento en el diseño, el tipo de agencia de los AMAs, las teorías éticas que incorporan –en el caso de que lo hagan–, y también nos informan sobre su desarrollo. Ahora bien, todos estos avances no están más que en sus fases iniciales y no se han validado fuera del laboratorio (Cervantes *et al.* 2020, 527). Asimismo, para estos autores los diferentes AMAs son agentes morales explícitos que distan mucho, al menos en el estado actual de su desarrollo, de poder equipararse con los agentes morales plenos. Los agentes morales explícitos se diferencian de los agentes implícitos y de los plenos. Los implícitos son capaces de actuar moralmente, pero sin haber sido programados para diferenciar las buenas y las malas conductas. Los agentes morales plenos hacen referencia a los seres humanos, con creencias, deseos, intenciones, libre voluntad y consciencia sobre sus acciones (Moor 2006, 19-21).

Esta diferencia entre agentes morales explícitos, implícitos y plenos ha suscitado una viva discusión en el terreno académico sobre el rol que deberían desempeñar estos AMAs en nuestras sociedades. En este sentido, han surgido voces críticas desde la ética, la computación y la robótica respecto de las virtudes de las máquinas éticas. Muchas de estas autoras y autores aducen que la moralidad debe quedar relegada al ámbito humano y que los agentes morales artificiales no pueden ni deben suplantar a los sujetos humanos (Bryson 2018; Fossa 2018; Johnson y Verdicchio 2018; Sharkey 2020; van Wynsberghe y Robbins 2019). Por eso, añaden, los robots, las inteligencias morales artificiales y las máquinas éticas deberían estar siempre bajo el control de los seres humanos (Bryson 2010). Solo a estos últimos correspondería responder por las responsabilidades de las acciones de los agentes artificiales creados de manera deliberada.

2.2. ¿Discontinuidad entre la agencia moral artificial y la humana?

La filosofía, sociología y antropología de la tecnología (Latour 2002; Verbeek 2014), la filosofía moral y la ética de la información (Floridi y Sanders 2004; Gunkel 2012, 2020) han debatido

ISSN 1989-7022

Dilemata, año 13 (2021), nº 34, 105-119

y reflexionado por extenso acerca de la posibilidad o imposibilidad de equiparar la agencia moral humana y la agencia moral artificial[3]. Por tanto, la consideración de la agencia tecnológica como agencia moral no se plantea solo en el contexto de los AMAs, sino en el ámbito más general de la filosofía de la tecnología.

Esto nos conduce a plantearnos, sucintamente, qué caracteriza a la agencia moral a partir sus elementos constituyentes y su naturaleza en un contexto en el que se plantea que la agencia moral está incorporada en las infotecnologías, los robots y los AMAs.

Tradicionalmente, la agencia moral plena se ha pensado desde criterios como la intencionalidad de actuar conforme a objetivos, valores o principios morales; la deliberación sobre esos principios; y la autonomía para decidir y actuar en un sentido o en otro. A los agentes morales se les atribuye responsabilidad sobre las consecuencias de sus acciones. Y, asimismo, se presupone que la agencia moral humana incorpora un componente de consciencia para poder justificar el porqué de nuestras elecciones y actuaciones (Himma 2009; Coeckelbergh 2020).

En lo que atañe a su funcionamiento, las infotecnologías son consideradas artefactos que incorporan una agencia moral y se denominan "ética sin mente" (*mindless morality)* o una "infraética" (Floridi y Sanders 2004; Floridi 2014, 2017)[4]. La agencia moral de las infotecnologías o de los AMAs se trabaja en términos operacionales y pierde elementos de la agencia moral humana como la consciencia o la responsabilidad del agente. Los agentes morales explícitos o la ética sin mente no incluyen consciencia ni deliberación sobre sus propias decisiones, tampoco emociones morales, y no pueden ser sujetos de atribución plena de responsabilidades por sus decisiones o acciones[5].

Una razón para que se considere el futuro potencial de los AMAs como mejor, es decir, superior a las debilidades de los razonamientos éticos de los seres humanos, estriba en poner de manifiesto que la traducción a los términos computacionales de las decisiones éticas pondría fin a las disputas sobre qué es la ética. El carácter artificial y computacional de los AMAs parecería ser mejor porque este permite superar las falibilidades del razonamiento moral del ser humano. En consonancia con esta idea, se sostiene que la traducción computacional de la moral pondría fin a las disputas entre la psicología y la filosofía moral (Wallach 2010, 249). En efecto, esta controversia entre las disciplinas atañe a la discusión sobre los mecanismos cognitivos y emocionales de los seres humanos para deliberar, evaluar y actuar moralmente, frente a la atención prestada por la filosofía no solo a las facultades morales, sino también a las teorías normativas, como el deontologismo, el consecuencialismo o la ética de la virtud, que entre sí son divergentes. Este hecho contribuye a que se cuestione el proyecto de la mejora moral porque no hay una comprensión previa de qué es la moralidad ni en qué consiste (Dubljević y Racine 2017; Racine *et al.* 2017; Tochibana 2017). En efecto, desde la filosofía, Douglas (2008) señala que cuando se habla de mejora de las disposiciones morales se puede estar haciendo referencia a la mejora de las emociones, a la mejora de las deliberaciones y de la racionalidad o a una mezcla de ambas (Douglas 2008, 231). Su propuesta, como se ha señalado, se centra en la mejora de las emociones, porque es en ese terreno donde encuentra razones para defender con plausibilidad que la atenuación de ciertas emociones contaría como una mejora moral con independencia de la teoría filosófica o psicológica que se defienda.

Ahora bien, la apuesta por una reproducción tecnológica de la agencia humana no parece solventar estos problemas, contrariamente a lo que piensan autores como Wallach (2010).

ISSN 1989-7022

Dilemata, año 13 (2021), nº 34, 105-119

Ante esta situación, cabría preguntarse si la reproducción computacional de la toma de decisiones morales no estaría agravando las disputas entre psicología, filosofía moral y teoría computacional respecto de qué es la moral. Y, es más, también es problemático plantear que la cesión del agente moral humano de su capacidad de toma de decisiones a los AMAs es buena para la ética sin más. Esto produciría una pérdida de autonomía del sujeto que, en vez de tratar de hacer frente a las deficiencias y falibilidades de la moral humana, optaría por una solución tecnológica más eficaz para hacer frente a los problemas morales. Estas cuestiones, por lo demás, no parecen ser abordadas lo suficiente desde el terreno de los defensores de las AMAs y, quizá, sí merecieran un mayor detenimiento.

Otra razón esgrimida generalmente para denostar el componente limitado cognitivamente y sesgado emocionalmente de la agencia moral humana –que no puede en la actualidad ser reproducida artificialmente– es que los AMAs superan en el proceso de toma de decisiones a los seres humanos y que las emociones no son necesarias para tomar las mejores decisiones morales (van Wynsberghe y Robbins 2019, 729).

En cualquier caso, desde la defensa de los AMAs se tiende a eliminar el componente emocional de la moralidad. Esto se produce en contraposición directa con el reclamo desde el proyecto de la biomejora neurofarmacológica de un mejor control de las emociones negativas y de un aumento de las emociones positivas, como la empatía y la compasión. Y, aunque se considere que los elementos racionales sean más importantes que los emocionales para la mejora moral, la maniobra de la eliminación de los componentes emocionales no supone una superación sin más de los procesos de toma de decisiones humanas por los agentes artificiales, sino una exhibición de las limitaciones de estos AMAs.

Desde posturas menos racionalistas, se aboga por la idea de que llegará un momento en que los AMAs puedan incorporar también un carácter sintiente dando lugar a una robótica afectiva (Lagrandeur 2015). Esto, no obstante, no es más que un proyecto especulativo, quizá irrealizable y que abre el campo a muchos nuevos interrogantes sobre la moralidad o inmoralidad del mismo (Bryson 2018; Johnson y Verdicchio 2018).

Ahora bien, la estrategia de suprimir todo aquello presente en la moralidad humana que no sea susceptible de ser traducido a términos computacionales y reclamar una superioridad en la toma de las decisiones los AMAs obvia elementos muy destacados. Uno de ellos es la consciencia del agente respecto de sus deliberaciones y la posibilidad, por ende, de responder por las consecuencias de sus acciones. El sujeto que actúa moralmente cuenta con la intención de hacerlo, puede atenuar o mitigar sus tendencias a decantarse por un curso de acción porque dichas tendencias estén en boga; también puede dialogar con sus emociones y evaluarlas racionalmente. Finalmente, puede reflexionar sobre el valor de usar sus capacidades racionales para intentar actuar bien, conforme a sus valores, con ayuda de información no refutada y con un análisis crítico sobre la plausibilidad de sus juicios morales, pese al reconocimiento de que las capacidades de la agencia moral humana son finitas y falibles (Lara y Deckers 2020, 283-285). Sin embargo, los agentes morales artificiales no pueden reflexionar sobre la importancia del papel del conocimiento y ni sobre la buena información en los procesos deliberativos; tampoco sobre cómo mejorar dicha deliberación y reflexión moral; y, por supuesto, no pueden ser autónomos, en el sentido de guiarse por lo que consideran más apropiado (Coekelberg 2020, 2054-2055). Como señalan Lara y Deckers (2020), estos agentes no tienen un sentido de la moralidad y no pueden hacer progresar la moralidad humana.

ISSN 1989-7022

Parece conveniente, pues, que se tome en serio la posibilidad de mejorar al agente moral humano sin sucumbir a un modelo exhaustivo, por el que se opte por una delegación plena a los AMAs para formular juicios morales, y se exploren otras vías que aborden directamente las falibilidades humanas con el fin de mejorarlas. Es aquí donde entra en juego la técnica del *neurofeedback*.

3. El neurofeedback: tecnología biomédica y digital para mejorar las deliberaciones morales

3.1. El neurofeedback como una técnica no invasiva de neuromodulación para la mejora moral

El *neurofeedback* es una interfaz cerebro-ordenador usada fundamentalmente por profesionales – terapeutas en su mayoría– por la que se enseña a las personas a que aprendan a modular las ondas de su cerebro. En el mercado, además, existen diferentes versiones portátiles de aparatos de *neurofeedback* para el entrenamiento en casa.

Las aplicaciones de esta interfaz en el ámbito terapéutico son muy variadas[6]. El *neurofeedback*, asimismo, es una técnica prometedora para investigar la relación entre la modulación cerebral y su relación con el comportamiento y la cognición (Enriquez-Geppert *et al.* 2017b*).* Y esta técnica se emplea en personas sanas para control de emociones (Zotev *et al.* 2014) y en la mejora de las funciones ejecutivas (Enriquez-Geppert *et al.* 2017a). Asimismo, el *neurofeedback* potencia la creatividad y el rendimiento en actividades como la música y la danza (Gruzelier 2014a y 2014b). Los buenos resultados de esta técnica se han aplicado también en el ámbito del deporte (Fronda *et al.* 2019). Y el neurofeedback ha despertado un creciente interés en el campo de la mejora moral (Darby y Pascual-Leone 2017; Nakazawa *et al.* 2016; Maslen y Savulescu 2016; Tachibana 2017a, 2017b, 2018).

Denominado en sus orígenes *biofeedback,* esta interfaz cerebro-ordenador permite que los sujetos aprendan a tener un control de sus ondas cerebrales. Estas se clasifican atendiendo a su frecuencia (Hz). En total, se diferencian cinco tipos de ondas, en función de sus ciclos por segundo, siendo estas de menor a mayor las siguientes: Delta, Beta, Alpha, Theta y Gamma (Hammond 2011)[7].

Gruzelier (2014a), por ejemplo, señala que trabajar las ondas Theta repercute en el aprendizaje, la memoria de trabajo, el procesamiento de información, la atención y la concentración. Las ondas Alpha están relacionadas con la atención, con el desempeño en la detección visual y con el razonamiento espacial (Enriquez-Geppert *et al.* 2017a, 150-153). Zotev *et al.* (2014) muestran como el trabajo de las ondas Gamma permite una reducción de las emociones negativas. Asimismo, el entrenamiento conjunto de las ondas Alpha y Theta se traduce en una mejora en la creatividad, de la práctica, y un aumento de las emociones agradables y placenteras asociadas a la experiencia artística de músicos y bailarines (Gruzelier 2014b).

Cabe señalar que el *neurofeedback* tiene un gran potencial en tres vertientes. La primera es la terapéutica, como se ha señalado. La segunda es la de la investigación de la relación de las ondas cerebrales con el comportamiento y la emoción. La tercera atañe a la mejora de personas sanas en ámbitos como la creatividad artística, la mejora cognitiva, la mejora en el rendimiento deportivo y la de la moralidad. El *neurofeedback* es una técnica en su infancia

Dilemata, año 13 (2021), nº 34, 105-119

ISSN 1989-7022

Dilemata, año 13 (2021), nº 34, 105-119

y, como señalan los investigadores de esta técnica, quedan aún muchas lagunas por cubrir[8]. Aun así, los prometedores resultados obtenidos dentro del laboratorio pueden verse refrendados también fuera del mismo con los usos domésticos de los dispositivos portátiles que son accesibles desde el mercado. En este sentido, los investigadores Enriquez-Geppert *et al.* (2017b) sostienen que la neuromodulación fuera del laboratorio tiene como consecuencia positiva la de validar los resultados obtenidos dentro de este. La propuesta de inclusión del *neurofeedback* para entrenar a los sujetos a neuromodularse y mejorar en sus deliberaciones morales podría servir como referente externo a los usos en el laboratorio.

Asimismo, una de las grandes ventajas del *neurofeedback* es que escapa a algunas críticas que han recibido otras técnicas de estimulación cerebral. Estas son objeto de estudio y de una viva discusión en el terreno de la mejora moral. En líneas generales, diferentes investigaciones se muestran escépticas respecto del potencial del uso de estas técnicas para mejorar la moral. Por ejemplo, Dubljević y Racine (2017, 346-348) concluyen la estimulación cerebral profunda afecta a los componentes emocionales de la moralidad, pero esta estimulación no afecta al componente deliberativo y de toma de decisiones morales. Además, señalan que técnicas no invasivas como la estimulación magnética transcraneal y la estimulación transcraneal por corriente continua no estimulan áreas cerebrales profundas y no tienen efectos en las emociones. Sin embargo, el *neurofeedback* escapa a algunas de estas críticas lanzadas contra estas técnicas biomédicas y de neuromodulación: frente a la estimulación transcraneal magnética y eléctrica, o la estimulación profunda del cerebro, el *neurofeedback* ha recibido menores críticas por sus efectos colaterales. Asimismo, el *neurofeedback* permite trabajar tanto la mejora de desempeños motores, cognitivos como el control de las emociones.

Por otro lado, como afirman Darby y Pacual-Leone (2017), ninguna técnica no invasiva (o invasiva) tiene el potencial de hacerse cargo de la complejidad y heterogeneidad del "cerebro moral". Estos autores ponen de relieve que el uso de técnicas no invasivas para estimular el cerebro repercute en procesos neuropsicológicos específicos que contribuyen a la conducta moral normal. Ahora bien, si estos procesos neuropsicológicos son alterados, podrían resultar en una mejora moral en ciertas situaciones, pero podrían conducir a conductas inmorales en otras. Por esta razón, concluyen que los objetivos del actual debate sobre la mejora moral son inalcanzables. Para Darby y Pacual-Leone, no existiría en principio una objeción en que las personas se mejorasen moralmente, si sus valores y creencias fuesen acordes a este proyecto, pues este proceso de mejora reforzaría la autonomía de los agentes. De hecho, estos autores proponen como objetivo modesto para la mejora moral que se mejoren las tendencias a actuar conforme a nuestras motivaciones morales. Pero, a su parecer, las técnicas de las que se dispone en la actualidad no permiten que haya ni mejora ni progreso moral.

Ahora bien, la interpretación del proyecto de la mejora moral de Darby y Pascual Leone (2017) es quizá poco acertada. Subyace a estos autores la idea de que la mejora tendría como fin el no actuar inmoralmente. No obstante, cabe objetar contra esta afirmación que existe una diferencia entre perfeccionarse, mejorarse o sobrepasar errores frecuentes de deliberación en el proceso de toma de decisiones morales, por un lado, y pensar que estos progresos suponen el fin de las deliberaciones y toma de decisiones erróneas, por otro. La mejora moral no es un proyecto que tienda a eliminar completamente las falibilidades humanas, sino a identificarlas, trabajar con ellas y tratar de superarlas.

ISSN 1989-7022

Dilemata, año 13 (2021), nº 34, 105-119

3.2. El control de las ondas cerebrales para la mejora moral

En el campo de la ética el *neurofeedback* se ha propuesto como una técnica para mejorar la moral atendiendo a tres aspectos: emocionales, comportamentales y racionales (Nakazawa *et al*. 2016; Maslen y Savulescu 2016; Tachibana 2017, 2018a y 2018b). Para Nakazawa y colaboradores (2016), el autocontrol de las emociones y del comportamiento que posibilita el entrenamiento del *neurofeedback* repercute en la mejora moral. Para estos autores, pues, la inhibición de respuestas emocionales negativas, la promoción de emociones positivas y el control del comportamiento son elementos claves de la moralidad, y todos estos elementos pueden trabajarse gracias al *neurofeedback*.

Maslen y Savulescu (2016) añaden que el entrenamiento con esta técnica favorece la deliberación moral. Para llegar a esta conclusión, los autores se basan en lo que en la propuesta de Nakazawa y colaboradores es un posible efecto colateral del *neurofeedback,* a saber, la irreversibilidad del aprendizaje adquirido como consecuencia de la plasticidad cerebral. Maslen y Savulescu consideran que este aprendizaje es fruto de la aceptación de las personas que se someten al entrenamiento neurocerebral. Por tanto, argumentan, el supuesto carácter irreversible del aprendizaje conseguido gracias al *neurofeedback* es positivo y deriva de la autonomía de la persona que decide neuromodularse para mejorar moralmente.

Para Tachibana (2018a), además, esta técnica debe comprenderse como una técnica tradicional de mejora de la moralidad en consonancia con la educación moral. Este autor defiende que gracias a esta técnica se podrían trabajar las disposiciones morales, bien sean estas emotivas o racionales, que se considerasen precisas.

La propuesta que se presenta aquí toma en consideración las aportaciones de estos estudios. Sin embargo, a diferencia de las anteriores consideraciones teóricas, plantea la mejora de las deliberaciones morales y, a fortiori, de la agencia moral humana desde, primero, un modelo de interacción plena que se oponga al modelo continuista. Y, segundo, desde una concepción de la mejora de la moral tecnológicamente mediada que plantea diferencias en cuanto a su forma y a su alcance con respecto a las técnicas tradicionales de la educación moral.

La apuesta por el *neurofeedback* como una técnica que contribuye a una mejora de la agencia moral desde una mediación tecnológica se presenta como un posible componente para una mejora moral de los aspectos racionales de la moralidad en línea con los objetivos del asistente socrático. En efecto, el modelo de interacción plena reclama que haya una mejora de los aspectos racionales de la moralidad. En este sentido, se habla de la necesidad de que el sujeto confronte con el asistente continuamente sus ideas, juicios y valores con la información empírica relevante y que pueda, por tanto, modificar sus creencias si estas no disponen de pruebas empíricas a su favor. En esta línea, el asistente socrático para la mejora moral habla, además, del requisito de la claridad conceptual y la mejora de la argumentación moral (Lara y Deckers 2020, 283-284).

Asimismo, una notable virtud del modelo de interacción plena es que vela por salvaguardar la pluralidad valorativa de los agentes morales, lo cual no está siempre garantizado por los AMAs, sobre todo si estos están programados desde una teoría normativa ética particular. Desde el modelo de la interacción plena se trata de que sean los propios agentes morales los que lleguen a sus conclusiones, sin que se presuponga una solución correcta a los problemas

ISSN 1989-7022

Dilemata, año 13 (2021), nº 34, 105-119

morales abordados. Este modelo aspira, pues, a que sean los propios agentes morales los que refuercen su autonomía. Este modelo de interacción plena, en síntesis, se contrapone al enfoque continuista y se asienta en la idea de que sí hay diferencias cualitativas entre los agentes morales humanos y las agencias morales artificiales. Sobre la base de estas diferencias, se pretende no tanto eliminar las falibilidades de la agencia humana, sino enmendarlas y, por tanto, abrir una vía para el progreso moral. En este contexto, las potencialidades del *neurofeedback* como técnica para la mejora de las deliberaciones morales se basa en la idea de que gracias al aprendizaje derivado de la neuromodulación, el agente moral se encontraría en un estado óptimo, desprovisto de distracciones, capaz de ser autocrítico con los sesgos presentes en sus juicios, o con las creencias no justificadas. En síntesis, el agente moral deliberaría mejor y ganaría en autonomía moral.

Por último, cabe destacar que la defensa del *neurofeedback* para mejorar la autonomía de la agencia moral humana escapa a algunas críticas de las que son objeto los proyectos de la mejora moral. Una de ellas es la supuesta merma de la autonomía, y otra la supuesta alteración de la identidad del sujeto (Specker *et al*. 2014).

Con respecto al primer riesgo, la idea de aprender a neuromodularse para deliberar mejor e interactuar de mejor manera con un asistente moral no supone, primero, una merma de la autonomía. Bien al contrario, se trataría de conseguir una potenciación de esta. En los casos de los AMAs, como acertadamente explican Lara y Deckers (2020), sí se encuentra un riesgo de perder autonomía al exigir los AMAs que los sujetos deleguen en ellos la capacidad de analizar los problemas morales y proponer soluciones para los mismos desde la perspectiva de la programación de la máquina.

Tampoco con esta propuesta se corre el riesgo de que se vea afectada la identidad del individuo. El riesgo de afectar a la identidad lo identifican Specker *et al*. (2014) con la técnica invasiva de la estimulación cerebral profunda. Sin embargo, como se comentó con anterioridad, esta última parece tener mayor repercusión en el sistema límbico y en la emoción que en los aspectos raciones de la toma de decisiones morales. La neuromodulación implica, por el contrario, un entrenamiento voluntario y las alteraciones de las ondas cerebrales. Sus repercusiones positivas para la deliberación moral pueden verse como resultado, justamente, de un consentido y paciente trabajo de los agentes morales (Maslen y Savulescu 2016). La plasticidad neurocerebral y, por tanto, la capacidad de las personas de aprender a neuromodularse dependen del éxito del entrenamiento en el que el sujeto participa. La apuesta por la mejora moral desde un modelo de interacción plena a través de un asistente moral socrático, del que forma parte el uso del *neurofeedback*, es una apuesta, asimismo, por el progreso moral. Esto, desafortunadamente, es altamente dudoso que se produzca con el uso de los AMAs (Lara y Deckers 2020, 279). Y en ningún caso, la idea de que puede conseguirse un progreso moral implicaría que el sujeto quedase determinado para actuar moralmente.

La mediación tecnológica para la mejora moral no supone, por tanto, que las personas queden alteradas o modificadas en su identidad, sino que puedan disponer de unos mecanismos para lograr su proyecto de ser mejores moralmente. Y en este proyecto, el *neurofeedback* podría colaborar mejorando la auto-reflexión, la consciencia y el control sobre la propia actividad cerebral de las personas, con las consecuentes repercusiones positivas para la deliberación moral.

ISSN 1989-7022

Dilemata, año 13 (2021), nº 34, 105-119

Para finalizar, el *neurofeedback* como técnica para la mejora de las deliberaciones morales presenta algunos inconvenientes que conviene mencionar. En primer lugar, la práctica del *neurofeedback*, aunque más segura e inocua que otras prácticas de neuromodulación, puede tener como consecuencia la aparición de algunos problemas como cansancio, somnolencia o dolor de cabeza (Hammond 2011). En segundo lugar, esta técnica requiere tiempo para que sea efectiva, pues requiere de sesiones de entrenamiento para el aprendizaje de la neuromodulación. Frente a las técnicas neurofarmacológicas y los AMAs, que actúan más rápidamente, la interfaz cerebro-ordenador demanda un proceso de aprendizaje que, posteriormente, debería completarse con la interacción con el asistente moral socrático. Esto puede ser visto como una gran desventaja en un mundo en el que exigimos cada vez más celeridad y respuestas inmediatas a nuestras demandas. Este inconveniente, sin embargo, atañe a un momento inicial: el aprendizaje y dominio de una técnica y/o el entrenamiento en una actividad (deportiva, cognitiva, musical, de razonamiento moral) forman parte de la adquisición de destrezas, dominio y mejora de la técnica en cuestión. Una vez dominadas las técnicas, el tiempo empleado es menor.

4. Conclusiones

El proyecto de mejora moral cuenta con varias vertientes para su posible desarrollo. En las líneas precedentes se han visto: la vía neurofarmacológica, centrada fundamentalmente en las emociones morales; la vía de los AMAs, centrada en los procesos de toma de decisiones morales desde una máquina ética; y la vía de la mejora de los aspectos racionales de la agencia humana a través de la mediación tecnológica.

En este artículo se ha hecho hincapié en los componentes racionales de la agencia moral humana. En este contexto, se han comparado los modelos continuista y de interacción plena. Y en esta comparación se ha puesto de manifiesto que existe una asimetría insalvable en lo que respecta a la agencia moral de los seres humanos y los AMAs. Así pues, con independencia de que en el futuro se desarrollen AMAs con una capacidad de decisión moral superior a la de los agentes morales humanos, hay un valor añadido en el hecho de querer mejorarse moralmente que no se podría lograr con la delegación plena de nuestra toma de decisiones morales a los agentes morales artificiales. En efecto, la consciencia de que podemos mejorar moralmente y la ganancia de autenticidad asociada a esta (Parens 2005) no pueden estar instanciadas en los AMAs. Las experiencias humanas que son afines a nuestras cualidades, intereses y valores –como practicar un deporte, tocar un instrumento o actuar éticamente– nos colman con experiencias positivas. Por esta razón, se defiende la mejora moral con un asistente moral interactivo, del que formaría parte el *neurofeedback* para mejorar nuestras deliberaciones morales. Desde este proyecto, por lo demás, no se pretende poner fin a las controversias de lo que es más importante en ética, ni sobre cuáles son los valores que debería movilizar un asistente moral, ni tampoco sobre cuál es el futuro de la ética. Los objetivos que se han planteado son, pues, más modestos que los que se persiguen desde la defensa de los AMAs.

La propuesta que se ha presentado, por lo demás, considera que la técnica de neuromodulación del *neurofeedback* puede ser un útil componente de un proyecto de mediación tecnológica de la mejora moral humana. Esto se realizaría en el marco de un proceso de interacción plena que, primero, rechazase la tesis continuista; segundo, reconociese las falibilidades de las facultades humanas; tercero tomase en consideración que toda técnica para la mejora

ISSN 1989-7022

Dilemata, año 13 (2021), nº 34, 105-119

moral no puede erigirse como una mejora total, sino parcial. Los motivos para actuar moralmente pueden diferir de las motivaciones para mejorarse moralmente a través de la mediación tecnológica y, en efecto, estos problemas no pueden ser resueltos por ningún asistente de Inteligencia Artificial (Lara y Deckers 2020, 284).

En cualquier caso, el *neurofeedback*, como parte del asistente socrático para la mejora moral, estaría enfocado a fortalecer los aspectos racionales de la moralidad, en concreto, mejoraría los aspectos deliberativos de la agencia moral humana. Y, aunque el componente deliberativo o racional de la moral no pueda por sí mismo ser concebido como lo definitorio o más importante de la agencia moral humana, cabría concluir que no por ello su valor puede ser desdeñado completamente.

Referencias bibliográficas

Agar, N. (2013). "Why is it possible to enhance moral status and why doing so is wrong?". *Journal of Medicine Ethics*, 39, pp. 67-74.

Brey, P. (2018). "The strategic role of technology in a good society". *Technology in Society,* 52, pp. 39-45.

Bryson, J. J. (2010). "Robots should be slaves", en Wilks, Y. (Ed.), *Close engagements with artificial companions: Key social, psychological, ethical and design issues.* Amsterdam, John Benjamins, pp. 63–74.

Bryson, J. J. (2018). "Patiency is not a virtue: the design of intelligent systems and systems of ethics". *Ethics and Information Technology*, 20, pp. 15–26.

Cervantes, J. A., López, Sonia y Rodríguez, L. F., Cervantes, S., Cervantes, F. Y Ramos, F. (2020). "Artificial Moral Agents: A Survey of the Current Status". *Science and Engineering Ethics.* 26, pp. 501–532. DOI: https://doi.org/10.1007/s11948-019-00151-x

Coeckelberhg, M. (2020). "Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability". *Science and Engineering Ethics,* 26, pp. 2051–2068. DOI: https://doi.org/10-1007/s11948-019-00146-8

Coeckelberhg, M. (2018). "Technology and the good society: A polemical essay on social ontology, political principles, and responsibility for technology". *Technology in Society,* 52, pp. 4-9. DOI: https://dx.doi.org/10.1016/j.techsoc.2016.12.002

Darby, R. R. & Pascual-Leone, A. (2017). "Moral Enhancement Using Non-invasive Brain Stimulation". *Frontiers in Human Neuroscience,* 11 (77). DOI: https://doi.org/10.3389/fnhum.2017.00077

De Melo-Martín, I. & Salles, A. (2015). "Moral Bioenhancement: Much Ado About Nothing". *Bioethics,* 29 (4), pp. 223–232. DOI: https://doi.org/10.1111/bioe.12100

DeGrazia, D. (2014). "Moral enhancement, freedom, and what we (should) value in moral behaviour". *Journal of Medical Ethics*, 40 (6), pp. 361-368.

Douglas, T. (2008). "Moral enhancement. *Journal of Applied Philosophy"*, 25 (3), pp. 228-245.

Douglas, T. (2013). "Moral Enhancement via Direct Emotion Modulation: A Reply to John Harris". *Bioethics,* 27 (3), pp. 160-168.

Dubljević, V. & Racine, E. (2017). "Moral Enhancement Meets Normative and Empirical Reality: Assessing the Practical Feasibility of Moral Enhancement Neurotechnologies". *Bioethics*, 31 (5), pp. 338–348. DOI: https://doi.org/10.1111/j.1467-8519.2011.01919.x

Enriquez-Geppert, S., Huster, R. J., Ros, R. J. & Wood, G. (2017a). "Neurofeedback*",* en Colzato L. (Ed.). *Theory-Driven Approaches to Cognitive Enhancement.* Cham, Switzerland, Springer, pp. 149-165. DOI: https://doi.org/10.1007/978-3-319-57505-6

Enriquez-Geppert, S. Huster R. J & Herrmann, C. S. (2017b). "EEG-Neurofeedback as a Tool to Modulate Cognition and Behavior: A Review Tutorial". *Frontiers in Human Neuroscience*, 11 (51). DOI: https://doi.org/10.3389/fnhum.2017.00051

Fossa, F. (2018). "Artificial Moral Agents: Moral mentors or sensible tools". *Ethics and Information Technology*, 20, pp. 115–126. DOI: https://doi.org/10-1007/s10676-018-9451-y

Floridi, L., & Sanders, J. W. (2004). "On the morality of artificial agents". *Minds and Machines*, 14 (3), pp. 349–379.

Floridi, L. (2014). "Artificial Agents and Their Moral Nature", en Kroes P. y Verbeek, P. P. (Eds.). *The Moral Status of Technical Artefacts, Philosophy of Engineering and Technology*. Heidelberg, New York, Dordrecht, London, Springer, pp. 185-212. DOI: http://doi.org/10.1007/978-94-007-7914-3_2.

Floridi, L. (2017). "Infraethics–on the Conditions of Possibility of Morality". *Philosophy of Technology*, 30, pp. 391–394. DOI: https://doi.org/10.1007/s13347-017-0291-1

Fronda, G., Crivelli, D., & Balconi, M. (2019). "Neurocognitive enhancement: Applications and ethical issues". *NeuroRegulation*, *6* (3), pp. 161–168. DOI: https://doi.org/10.15540/nr.6.3.161

Gruzelier, J. H. (2014a). "EEG-neurofeedback for optimising performance. I: A review of cognitive and affective outcome in healthy participants". *Neuroscience and Biobehavioral Reviews,* 44, pp. 124-141. DOI: http://doi:10.1016/j.neubiorev.2013.09.015

Gruzelier, J. H. (2014b). "EEG-neurofeedback for optimising performance II: Creativity, the performing arts and ecological validity". *Neuroscience and Biobehavioral Reviews,* 44, pp. 142-158. DOI: https://doi.org/10.1016/j.neubiorev.2013.11.004

Gunkel, D. J. (2012). The machine question: *critical perspectives on AI, robots, and ethics*. Cambridge, MIT Press.

Gunkel, D. J. (2020). "Mind the gap: responsible robotics and the problem of responsibility". *Ethics and Information Technology*, 22, pp. 307–320. DOI: https://doi.org/10.1007/s10676-017-9428-2

Hammond, D. C. (2011). "What is Neurofeedback?: An Update". *Journal of Neurotherapy*, 15, pp. 305–336. DOI: https://doi.org/10.1080/10874208.2011.623090

Harris, J. (2013a). "Ethics is for bad guys! Putting the 'Moral' into Moral Enhancement". *Bioehics,* 27 (3), pp. 169–173. DOI: https://doi.org/10.1111/j.1467-8519.2011.01946.x

Harris, J. (2013b). "Moral Progress and Moral Enhancement*", Bioethics* 27 (5), pp. 285–290. DOI: https://doi.org/10.1111/j.1467-8519.2012.01965.x

Hauskeller, M. (2013). *Better humans? Understanding the enhancement project*. Durham, Acumen.

Himma, K. E. (2009). "Artificial agency, consciousness, and the criteria for moral agency: what properties must an artificial agent have to be a moral agent?" *Ethics and Information Technology*, 11, pp. 19–29. DOI: https://doi.org/10.1007/s10676-008-9167-5

Johnson, D. & Verdicchio, M. (2018). "Why robots should not be treated as animals". *Ethics and Information Technology*, 20, pp. 291–301. DOI: https://doi.org/10.1007/s10676-018-9481-5

Lagrandeur, K. (2015). "Emotion, Artificial Intelligence, and Ethics", en Romportl, J. Zackonova, E. y Kelemen, J. (Eds.) *Beyond Artificial Intelligence. The disappearing of Human-Machine Divide*. Cham Heidelberg New York Dordrecht London, Springer, pp. 97-110. DOI: https://doi.org/10.1007/978-3-319-09668-1

Lara, F. & Deckers, J. (2020). "Artificial Intelligence as a Socratic Assistant for Moral Enhancement". *Neuroethics* 13, pp. 275–287. DOI: https://doi.org/10/s12152-019-09401-y

Latour, B. (2002). "Morality and Technology. The End of Means*". Theory, Culture and Society*, 19(6), pp. 247-260.

Maslen, H. & Savulescu, J. (2016). "Neurofeedback for Moral Enhancement: Irreversibility, Freedom, and Advantages Over Drugs". AJOB Neuroscience 7 (2), pp. 120-122. DOI: https://doi.org/10.1080/21507740.2016.1189976

Moor, J. H. (2006). "The nature, importance, and difficulty of machine ethics". *IEEE Intelligent Systems,* 21(4), pp. 18–21.

ISSN 1989-7022

ISSN 1989-7022

Dilemata, año 13 (2021), nº 34, 105-119

Nakazawa, E., Yamamoto, K., Tachibana, K., Toda, S. Takimoto, Y., y Akabayashi, A. (2016). Ethics of decoden neurofeedback in clinical research, treatment, and moral enhancement. *AJOB Neuroscience,*7 (2), pp. 110-117. DOI: https://doi.org/10.1080/21507740.2016.1172134.

Parens, E. (2005). "Authenticity and Ambivalence: Toward Understanding the Enhancement Debate". *Hastings Center Report*, 35 (3), pp. 34-41.

Persson, I. & Savulescu, J. ***(2014). Unfit for the Future***: *The Need for Moral Enhancement.* Oxford, Oxford University Press.

Racine, E., Dublejevic´, V., Jox, R. J., Baertschi, B., Christensen, J. F., Farisco, M. Jotterand, F., Kahane, G. & Müller, S. (2017). "Can neuroscience contribute to practical ethics? A critical review and discussion of the methodological and translational challenges of the neuroscience of ethics". *Bioethics,* 31 (5), pp. 328–337. DOI: https://doi.org/10.1111/bioe.12357

Sandel, M. (2009). "The Case Against Perfection: What's Wrong with Designer Children, Bionic Athletes, and Genetic Engineering", en Savulescu, J. y Bostrom, N. (Eds.) *Human Enhancement.* Oxford, Oxford University Press, pp. 71-89.

Savulescu, J. & Maslen, H. (2015). "Moral enhancement and Artificial Intelligence: Moral AI?", en Romportl, J. Zackonova, E. y Kelemen, J. (Eds.) *Beyond Artificial Intelligence. The disappearing of Human-Machine Divide.* Cham Heidelberg New York Dordrecht London, Springer, pp. 79-95. DOI: https://doi.org/10.1007/978-3-319-09668-1

Sharkey, A. (2020). "Can we program or train robots to be good?". *Ethics and Information Technology*, 22, pp. 283–295.

Specker, J. Focquaert, F., Raus, K., Sterckx, S & Schermer, M. (2014). "The ethical desirability of moral bioenhancement: a review of reasons". *BMC Medical Ethics*, 15 (67).

Tachibana, K. (2017). "Neurofeedback-Based Moral Enhancement and the Notion of Morality". *The Annals of the University of Bucharest - Philosophy Series,* 66 (2), pp. 25-41.

Tachibana, K. (2018a). "Neurofeedback-Based Moral Enhancement and Traditional Moral Education". *Humana Mente,* 11 (33), 19-42.

Tachibana, K. (2018b). "The Dual Application of Neurofeedback Technique and the Blurred Lines Between the Mental, the Social, and the Moral". *Journal of Cognitive Enhancement*, 2, pp. 397–403. DOI: https://doi.org/10.1007/s41465-018-0112-1

van Wynsberghe, A. & Robbins, S. (2018). Critiquing the reasons for making artificial moral agents. *Science and Engineering Ethics*, 25 (3), pp. 719-735. DOI: https://doi.org/10.1007/s11948-018-0030-8

Verbeek, P.P (2014). Some Misunderstandings About the Moral Significance of Technology (2014), en Kroes, P. and Verbeek, P.P. (Eds.). *The Moral Status of Technical Artifacts, Philosophy of Engineering and Technology,* 17, Heidelberg, New York, Dordrecht, London, Springer, pp. 75-88. DOI: https://doi.org/10.1007/978-94-007-7914-3_5,

Wallach, W. (2010). "Robot minds and human ethics: The need for a comprehensive model of moral decision making". *Ethics and Information Technology*, 12 (3), pp. 243–250.

Zotev, V. Phillips, R., Yuan, H., Misaki, M. & Bodurka, J. (2014). "Self-regulation of human brain activity using simultaneous real-time fMRI and EEG neurofeedback". *NeuroImage,* 85, pp. 985-995.

Notas

1 La mejora moral, como toda la reflexión sobre la mejora humana, se ha abordado en el terreno académico desde dos enfoques bien diferenciados: el primero es bioconservador, el cual rechaza de pleno toda técnica que tenga como fin mejorar las capacidades humanas y/o crear seres humanos mejorados con capacidades físicas y cognitivas que trasciendan a los seres humanos (convirtiéndose en poshumanos) (Sandel 2009). Frente a esta posición bioconservadora, el poshumanismo considera permisible y, en ciertos casos, necesarias las mejoras de las capacidades humanas con técnicas biomédicas y otras tecnologías. El proyecto de mejora moral dentro del poshumanismo, asimismo, se estudia atendiendo a técnicas que permiten la

ISSN 1989-7022

Dilemata, año 13 (2021), nº 34, 105-119

mejora de las disposiciones morales gracias a la ciencia y a la tecnología, y también reflexiona sobre el estatus moral de las personas mejoradas, lo cual no ha estado exento de controversia. Para Agar (2013), por ejemplo, en el caso de que hubiese poshumanos, estos tratarían a los seres humanos exactamente como se merecen, y les impondrían sacrificios legítimos desde su estatus superior, como los seres humanos imponen a otros seres vivos. Por esta razón, la apuesta de Agar pasa por oponerse a la creación de poshumanos con un estatus moral superior (2013, 72-73).

2 Por limitaciones de espacio, desafortunadamente, no se aborda en profundidad el tipo de agencia moral que incardinan estos AMAs, ni tampoco las argumentaciones a favor de que los AMAs cuenten con derechos y responsabilidades en un plano axiológico, ontológico, ético y político (Gunkel 2020).

3 Esta discusión, muy viva en el terreno de la filosofía de la tecnología, toma como elemento central la tesis del instrumentalismo. Esta tesis afecta al modo de comprender a los artefactos tecnológicos como meros instrumentos o como agentes artificiales que interactúan con agentes humanos (Gunkel 2012, 2020). Desde el rechazo del instrumentalismo se defiende que el programa de diseño, guion, o plan que incorporan los artefactos perfila, aunque de manera no determinista, las funciones y usos de las tecnologías. Los artefactos incorporan en su diseño su relación prevista con los usuarios. Un asistente artificial ético estaría diseñado para que las personas deliberasen mejor en el plano moral, no para otros fines. Estas interrelaciones entre artefactos y humanos moldean la experiencia humana y de la sociedad. Por eso se atribuye un rol más activo a las agencias artificiales desde el rechazo del instrumentalismo. La postura contraria afirmaría que toda tecnología no sería más que un mero medio para conseguir un fin, es decir, un instrumento. A estos no se les presupone ninguna identidad equiparable los agentes humanos, quienes deciden cómo se crean, los usos que pueden tener y cuándo no deberían ser utilizados. Desde esta perspectiva, la tesis continuista carece de sentido.

4 Los conceptos de "ética sin mente" o de "infraética" hacen referencia al tipo de agencia moral artificial de las infotecnologías. Todas las tecnologías digitales incorporan, a juicio de Floridi y Sanders (2004), una suerte de agencia moral que no es reflexiva, pero que se caracteriza por desempeñar un rol moral. Las responsabilidades éticas derivadas de las interrelaciones entre los agentes humanos y las tecnologías digitales exigiría una concepción nueva de la responsabilidad moral que esté compartida entre agentes humanos y artificiales. Para plantear esta posibilidad, los autores insisten en que la agencia moral de estas tecnologías se caracteriza por: la interactividad del agente con su entorno; la autonomía, entendida un cambio de estado de una máquina propiciado porque la máquina dispone de más de un estado; y por último la adaptabilidad, consistente en que el agente pueda cambiar las reglas en sus interacciones con el entorno. Estos elementos vendrían a ser los correlatos de la intencionalidad, la autonomía y la toma de decisiones de la agencia moral humana (Floridi 2104, 193-194).

5 La tesis del instrumentalismo y la tesis de la atribución de responsabilidades compartidas entre agencias humanas y no humanas avivan un apasionante debate. Lamentablemente, por cuestiones de espacio no podrán tampoco ser abordados estos temas.

6 Véase Hammond (2011) y Kober *et al.* (2015). El *neurofeedback* se usa para la recuperación de accidentes cerebrovasculares y problemas sensorio-motrices. Se emplea también como tratamiento para pacientes con TDAH, epilepsia, insomnio, fibromialgia, trastorno de estrés postraumático, trastorno del espectro autista y las adicciones, entre otros.

7 Las ondas Delta son muy lentas y de mucha amplitud; actúan en el sueño profundo y restaurador y también cuando nos sentimos adormilados. Las Theta son las que prevalecen cuando estamos soñando despiertos; se asocian con falta de eficacia mental y con inatención a la realidad exterior. En niveles muy lentos producen estados de relax; estás ondas están presentes en los momentos de vigilia previos al sueño. Las ondas Alpha se asocian con niveles de relajación. Si se cierran los ojos y se piensa en una imagen apacible, las ondas Alpha comienzan a aumentar. Las ondas Beta son ondas más rápidas y se asocian a estados mentales de actividad intelectual y de concentración. Las ondas Gamma son muy rápidas y se relacionan con la atención focalizada; estas ondas ayudan al cerebro a procesar y unir información de diferentes partes del cerebro (Hammond 2011).

8 Por ejemplo se desconocen cuáles son los mecanismos específicos responsables la plasticidad de las ondas y hay mucho trabajo que recorrer en el terreno del diseño de protocolos específicos para comparar los resultados de esta técnica con diferentes protocolos y a diferentes grupos.