

# LA CULPABILIDAD MÁS ALLÁ DE LA NEUROCIENCIA: ¿QUÉ SIGUE?

○ Andrés Felipe Díaz Arana\*

\*Abogado y filósofo por la Universidad de los Andes. Doctorando en Derecho en la Universidad Pompeu Fabra (España).  
Contacto: [af.diaz226@uniandes.edu.co](mailto:af.diaz226@uniandes.edu.co)

## PALABRAS CLAVE *KEYWORDS*

○ **Neurociencia**

*Neuroscience*

○ **Culpabilidad**

*Culpability*

○ **Inteligencia artificial**

*Artificial intelligence*

○ **Personalización de Internet**

*Personalization on the Internet*

○ **Cámaras de eco**

*Echo chambers*

**Resumen.** El debate de las neurociencias ha copado buena parte de la discusión sobre la culpabilidad jurídico-penal de las últimas décadas. Para algunos, la cuestión ya fue superada; para otros, sigue vigente. En cualquier caso, esta década nos arrastra inevitablemente a una nueva realidad que obliga a reflexionar sobre los retos que vienen para el *ius puniendi*. Mentes sin cerebros bioquímicos, inteligencia (y agencia) artificial, “cámaras de eco” virtuales y demás fenómenos contemporáneos y futuros pintan un panorama interesante para la controversia. Este artículo recoge el estado del debate en torno a las neurociencias e introduce algunos de los asuntos de los que la discusión debería ocuparse en el futuro.

**Abstract.** The criminal discussion about culpability has approached the neuroscience debate. For some authors, the question has already been solved. For others, it remains current. This decade inevitably drags us into a new reality, forcing us to reflect on the challenges of criminal law for years to come. Minds without biochemical brains, artificial intelligence (and agency), virtual echo chambers, and other contemporary and future phenomena paint an interesting outlook for the legal-criminal controversy. The starting point of this article is the actual state of the debate concerning neuroscience. It will introduce, as well, some of the issues that we consider should be discussed in the future.

Fecha de recepción: 20 de julio de 2020

Fecha de aceptación: 6 de agosto de 2020

## SUMARIO:

**I. Introducción. II. El pasado reciente de la discusión: neurociencias y filosofía de la mente. III. Inteligencia y agencia artificial: libertad más allá de las neurociencias. IV. Personalización de internet, “cámaras de eco” y acceso normal a la prohibición: desafíos y perspectiva a futuro. V. Reflexión final. VI. Fuentes de consulta.**

---

## I. INTRODUCCIÓN

En lo que va del siglo, buena parte de la discusión sobre la culpabilidad se ha ido en determinar el sentido y alcance que deben tener los avances de las neurociencias y, en particular, uno de ellos en la teoría material del delito y en la política criminal. Con algunas décadas de atraso respecto de su publicación original, los resultados de algún experimento encendieron el debate en nuestra tradición continental, por la aparente amenaza que representaban para los fundamentos sobre los cuales se erige, tradicionalmente, el sistema de imputación de responsabilidad jurídica (penal). Los científicos habrían dado con un descubrimiento que sacudiría las bases de la convivencia social.

A decir verdad, la idea de que la sociedad se enfrenta a una “nueva realidad” que amenaza con acabarla es, literalmente, tan antigua como la historia.<sup>1</sup> Aunque apenas se haya acogido en el derecho penal el término “sociedad del riesgo”,<sup>2</sup> la idea de que el riesgo es inherente a la vida en sociedad es antigua.<sup>3</sup> De hecho, hoy la doctrina mayoritaria reconoce que los bienes de la sociedad están expuestos constantemente al peligro de su destrucción y, en ese contexto, el derecho penal ha encontrado su sitio como administrador en vez de opositor del riesgo (Del Rosal Blasco, 2009) en un estado de cambio permanente.<sup>4</sup>

Sin embargo, cada tanto ocurre algún evento extraordinario que obliga a alterar la cadencia de la discusión y redirigir el debate hacia algunos viejos conceptos que se creían superados, a abrir algunas discusiones que

---

<sup>1</sup> Ciertamente, ya en el año 2,800 a.C., en una tablilla de la antigua Mesopotamia, un preocupado acadio escribía: “Hemos caído en tiempos perversos/El mundo se ha vuelto viejo y despiadado/Las políticas son corruptas/Y los hijos ya no respetan a los padres”. Tablilla cuneiforme 456 del imperio acadio, atribuida frecuentemente al Rey Naram-Sim de Caldea. Patrick (1913: 493) mencionó que: “We have fallen upon evil times, and the world has waxed very old and wicked, politics are very corrupt, children are no longer respectful of their parents”.

<sup>2</sup> Término que suele asociarse a la obra del sociólogo alemán Ulrich (1998) en la década de 1980.

<sup>3</sup> Al respecto, Hassemer (1992).

<sup>4</sup> Como lo apuntaron Hegel y, mucho tiempo antes, Heráclito.

se consideraban pasadas de moda. Tal fue el caso de los hallazgos de Libet (1983) en relación con la discusión sobre el libre albedrío en el derecho penal. Y, aunque para algunos ya ha sido superada, parece ser que desarrollos recientes en otras disciplinas abren dimensiones todavía más interesantes sobre esta cuestión.

En lo que sigue, ofrezco una revisión actualizada del estado de la discusión respecto de tres cuestiones íntimamente relacionadas con el libre albedrío. En primer lugar, me referiré al asunto de las neurociencias, que ocupó buena parte de la doctrina sobre culpabilidad jurídico-penal de la tradición continental en la última década. Hecho esto, me ocuparé de discusiones que, actualmente, se abren paso en relación con recientes avances en materia de inteligencia artificial. Finalmente, destinaré un espacio a presentar un asunto que prácticamente no ha sido explorado desde el campo específico del derecho penal y, en cambio, sí en otras disciplinas: la interferencia al acceso normal a la prohibición por sesgos cognitivos asociados a las cámaras de eco y la personalización de internet.

---

## II. EL PASADO RECIENTE DE LA CUESTIÓN: NEUROCIENCIAS Y FILOSOFÍA DE LA MENTE

Sobre este particular, conviene empezar por reconocer que, quizás por el revuelo que causaron algunos avances de las neurociencias en la discusión jurídico-penal de la última década sobre el determinismo,<sup>5</sup> aún sigue resonando la cuestión cada vez que se refiere al libre albedrío. De hecho, en el campo específico de la interacción entre neurociencias, derecho penal y filosofía de la mente,<sup>6</sup> la discusión no parece haber acabado del todo.

---

<sup>5</sup> En nuestra tradición continental, por ejemplo: Feijoo Sánchez (2011, 2012 y 2013), Demetrio Crespo (2014), Vives Antón (2013) y Pérez Manzano (2013). Entre las publicaciones más relevantes de la cuestión en la doctrina anglosajona: Jones, O. *et al.* (2011, 2013), Ginther *et al.* (2014), Morse, S y Roskies, A. (2013), Freeman, M. y Goodenough, O. (2009). En Colombia, Gómez Pavajeau, C. (2018).

<sup>6</sup> Hasta hace apenas unos años, no existían muchos escritos en nuestra tradición (aunque sí en la anglosajona) que se ocuparan de este asunto desde la interacción entre la filosofía de la mente y el derecho penal. No obstante, en España, Demetrio Crespo (2014: 111-124), pese a que no lo desarrolló con profundidad en ese entonces, ya advertía que “de la misma manera que sería equivocado probablemente pasar por alto los avances producidos en el ámbito de la física contemporánea, lo mismo sería predicible respecto a la filosofía de la mente”.

A decir verdad, en los últimos años, las publicaciones han seguido saliendo, aunque con menor frecuencia que antes<sup>7</sup> y, en esencia, con las mismas posturas emblemáticas enfrentadas. Veamos.

De un lado, siguen apareciendo posturas de corte realista que afirman la existencia del libre albedrío como un fenómeno ontológico de orden inmaterial, que gobierna y caracteriza a la actuación humana. Uno de los ejemplos recientes más notorios es el de Alonso Álamo (2018), quien, a la par que destaca la relevancia actual de la discusión en torno a las neurociencias y la inevitabilidad de que, para superar sus retos, se acuda a la filosofía de la mente,<sup>8</sup> adopta de fondo una postura según la cual, el libre albedrío tiene existencia real en un orden “enigmático”<sup>9</sup> inaccesible para las neurociencias, con fundamento en el cual es “preferible científicamente”<sup>10</sup> afirmarlo que negarlo.

De otro lado, también recientemente, y con apoyo en la filosofía de la mente, pero desde una perspectiva radicalmente distinta, Sánchez Vilanova (2019) se ha referido a la cuestión en un trabajo monográfico que presenta al “neuroderecho” como una nueva disciplina. En su opinión,

[e]l recurrente recurso al libre albedrío [...] arrastra el dualismo que todavía impera en nuestra sociedad, pero que el Derecho penal abandonó hace décadas. De hecho, parece que el clásico salto a la mente inmaterial nunca se produjo, dado que, al fin y al cabo, la libertad que fundamenta nuestra sociedad consiste en una capacidad explicable a partir

<sup>7</sup> Me refiero a la doctrina, escenario natural de la discusión, y en el cual se ha desarrollado casi en su totalidad. En la jurisprudencia no puede decirse que haya disminuido la producción al respecto; de entrada, porque antes no era significativa. En este escenario, algunas providencias se refieren tímidamente al particular. Pero, en comparación con la doctrina, la discusión de fondo sigue siendo prácticamente inexistente. En la ocasión más reciente, el Tribunal Supremo de España (2020) se pronunció sobre esta cuestión en los siguientes términos: “El papel de la neurociencia ha resucitado en la dogmática penal el interés por ciertos perfiles criminológicos que, en una visión histórica que se creía ya superada, predisponían al delito. Hoy asistimos a lo que con acierto se ha llamado una ‘revolución neurocientífica’. Las neurociencias, valiéndose de renovados métodos de experimentación y neuroimagen, están permitiendo un análisis de la mente humana inimaginable hasta hace bien poco. Las tesis defendidas por algunos neurocientíficos, que niegan toda escisión entre la mente y el cerebro, hasta el punto de que la actuación consciente no sería sino una ínfima expresión de nuestra actividad cerebral, han servido para reabrir una controversia histórica, a saber, el determinismo como explicación de muchos de nuestros actos. Se ha dicho, por entusiastas defensores de esta perspectiva, que ‘no hacemos lo que decidimos, sino que decidimos lo que vamos a hacer de todas maneras’” (p. 10). Sin embargo, y quizás de manera inesperada, dado el hilo que empezaba a abrirse paso, acto seguido descarta el asunto, alegando: “Es cierto que estamos ante un debate todavía en ciernes y sobre el que la Sala no tiene necesidad de pronunciarse”. (STS, 814/2020, 05.05.2020, p. 11)

<sup>8</sup> “Los avances de las neurociencias han reavivado una discusión que parecía hace tiempo superada’ y que ‘los nuevos conocimientos científicos y los interrogantes que abren permiten vislumbrar un futuro en el que no va a ser posible fundamentar la culpabilidad jurídica de espaldas a la ciencia, ni tampoco de la filosofía de la mente’”. (Álamo, 2018: 6-7)

<sup>9</sup> Considera Alonso Álamo (2018: 26-27) que “desde la realidad de la experiencia individual única se transita hacia una existencia real aunque enigmática. Pues [sic] las bases de un objeto explicable científicamente, como la conciencia, no necesariamente son susceptibles de estudio científico”.

<sup>10</sup> “El reconocimiento de una existencia real aunque enigmática de la libertad de voluntad es una opción científicamente preferible a la de su negación” (Álamo, 2018: 27).

de la evolución humana, y la Neurociencia, más que negarla, la valida (Sánchez Villanova, 2019: 174-175).

Igualmente, en el contexto de la específica intersección entre el derecho penal y la filosofía de la mente en la discusión sobre las neurociencias, Díaz Chunga (2020) ha tomado partida. Desde una perspectiva, otra vez, distinta de las anteriores, este autor acude al funcionalismo sistémico para defender una forma de “libertarismo” desde la cual afirma la existencia de espacios de indeterminación en los procesos de toma de decisión de las personas.<sup>11</sup> Sin perjuicio de lo anterior, ofrece una reflexión interesante acerca del espacio que ocupan las neurociencias en la discusión sobre el determinismo: “[a]ctualmente las neurociencias [...] han podido reabrir el debate y la discusión sobre este tema [libre albedrío], inclusive nos han recordado que existen posiciones plausibles que niegan el libre albedrío sin la necesidad de recurrir a información neurobiológica” (Díaz Chunga, 2020: 221).

Esto introduce la muy importante idea de que, quizás, la cuestión sobre el libre albedrío no dependa de los avances en neurociencias.<sup>12</sup> En verdad, parece ser que la discusión sobre el fundamento de la culpabilidad comprende mucho más que la muy limitada discusión en torno a las implicaciones de algunos avances neurocientíficos para el derecho penal. En este sentido, coincido con Nieva Fenoll en cuanto a que, en el estado actual de la discusión, parece ya haberse superado la “amenaza”<sup>13</sup> que algunos identificaron en algún momento.<sup>14</sup>

<sup>11</sup> Díaz Chunga (2020: 236) explica que “nos encontramos comprometidos con defender alguna forma de libertarismo. Es decir, con la posición según la cual las personas tienen libre albedrío y, en consecuencia, algunas de sus acciones no se encuentran determinadas, contradiciendo al menos parcialmente el determinismo del mundo”.

<sup>12</sup> En efecto, la imposibilidad de obrar en contrario o, correlativamente, la necesidad de asumir que el agente no pudo haber obrado de modo distinto, no tiene por qué dirigir la discusión hacia el debate en torno a las neurociencias. En su momento, esta fue la postura que adopté sobre el particular (Díaz Arana, 2016). Creo que los recientes avances tecnológicos y, en general, de la discusión, nos han llevado a nuevas dimensiones en las cuales ahora resulta menos extravagante hablar de “mentes sin cerebro (bioquímico)”, como se explica más adelante.

<sup>13</sup> En palabras de Vives Antón (2002: 232), la amenaza de las conclusiones de algunos neurocientíficos no se reducía a la culpabilidad, sino que “todo el lenguaje de la acción quedaría deslegitimado por ella”. De acuerdo con Sánchez Villanova (2019: 138), “más que afectar a la fundamentación de la culpabilidad, descubrimientos tan populares como el de Libet, incidirían, como previamente apuntamos, en la existencia misma de la acción”.

<sup>14</sup> Para Nieva Fenoll (2016: 120), “Benjamin Libet no inició los estudios sobre neurociencia, sino que son muy anteriores. Ni siquiera puede decirse, ya a día de hoy, que sus conclusiones posean importancia alguna en el estudio del Derecho pese al revuelo que generaron no hace tanto”. Aunque coincido en que hoy en día las conclusiones de Libet no deberían seguir suscitando problema alguno, creo que no por ello los resultados del debate dejan de ser importantes. Muchos temas en el derecho penal han quedado ya superados, pero siguen siendo importantes como peldaños de la escalera. Ciertamente, el debate intenso que suscitó la cuestión (y que no puede decirse que haya desaparecido del todo) arrojó notables avances en la dogmática penal, y descartar sus conclusiones solamente porque la cuestión haya quedado agotada no hace justicia a su relevancia histórica.

Sin embargo, considero que no por ello la cuestión ha perdido relevancia. Al respecto, es preciso aclarar que la discusión sobre determinismo y responsabilidad sigue avanzando (y seguramente nunca dejará de hacerlo), pero el punto específico sobre la relación entre las neurociencias y el derecho penal, en mi opinión, ha rendido ya su capacidad.<sup>15</sup> Nos ha mostrado (¿o recordado?) que la libertad sobre la cual se erige el sistema jurídico-penal de atribución de responsabilidad no *depende* de la comprensión de la “circuitaría de la mente”, en los términos de Damasio (2006), sino que se nutre de los avances de este, así como de todos los demás frentes.

El que esta discusión haya servido para aceptar esto es muy valioso en sí y, por eso, constituye un referente importante en cualquier texto que aborde la materia. Pero el debate en torno a la autodeterminación no debe reducirse a la cuestión acerca del alcance de las conclusiones de Libet (ni, para estos efectos, de las neurociencias en general). Por el contrario, debe abarcar puntos tradicionales que siguen tan vigentes como siempre y otras nuevas dimensiones que, con seguridad, orientarán el debate jurídico-penal hacia el futuro.

De hecho, valiosas publicaciones han salido recientemente en el espacio de encuentro entre derecho (en general, no específicamente penal) y filosofía (de la mente y acción), sobre temas que nada tienen que ver con procesos bioquímicos ni cerebros humanos. Tal es el caso, por ejemplo, de la inteligencia y agencia artificial.

---

### III. INTELIGENCIA Y AGENCIA ARTIFICIAL: LIBERTAD MÁS ALLÁ DE LAS NEUROCIENCIAS

En una reciente publicación dedicada, en parte, a la llamada “LegalTech” que, dicho sea de paso, según Dolm (2019) es un mercado económico que creció 713% el año pasado, Delgado de Molina Rius (2020) recoge los orígenes e historia de la inteligencia artificial (IA) para explicar su potencial actual para el derecho, al tiempo que Sánchez del Campo (2020) actualiza la discusión en torno a la “agencia artificial”. En la primera de las publicaciones referidas, el autor ofrece una actualización en la historia de la inteligencia artificial, desde el conocido nacimiento del término en 1956 en el

---

<sup>15</sup> Para Feijoo Sánchez (2011: 9), “[n]o estamos más que ante un nuevo capítulo de un folletín que empezó hace mucho tiempo. Las neurociencias ofrecen una nueva base científica para el debate, pero las cuestiones de fondo siguen siendo las mismas”.

marco de la conferencia de Dartmouth College, Nuevo Hampshire, hasta las novedades más relevantes en la época contemporánea. Sobre el estado actual de la cuestión, presenta interesantes consideraciones como que “el 90% de los datos existentes se han generado en los dos últimos años” (p. 103) (en relación con el campo bautizado como “Big Data”), que “la capacidad procesadora de los ordenadores ha aumentado exponencialmente desde 1993, tal y como predecía la Ley de Moore” (p. 104) (algo que, de hecho, es reiterado con frecuencia en escritos que se refieren al asunto)<sup>16</sup> y que, hoy en día, las máquinas son capaces de reconocer su entorno y, autónomamente, orientar su comportamiento en concordancia —en alusión a los interesantes avances en *deep learning*—. <sup>17</sup>

Desde la filosofía de la mente, esta es una idea muy atractiva. Los llamados “modelos computacionales de la mente” suelen partir de una perspectiva funcional de la mente, según la cual “la naturaleza de un estado mental es justo como la naturaleza de un autómatas: constituida por sus relaciones con otros estados y con entradas y salidas” (Block, 1995: 324). Lo interesante de esta propuesta es que considera que la mente no es *como* una máquina, sino que *es* una máquina (Smolensky, 1994).

A decir verdad, “inteligencia artificial” es un término que adopta casi tantas definiciones como autores que escriben sobre el particular (Legg y Hutter, 2007). Sin embargo, quizás una de las más aceptadas —¿por su vaguedad?— sea aquella ofrecida por Marvin Minsky (1968: 6), fundador del laboratorio de Inteligencia Artificial del MIT, según la cual inteligencia artificial es “la ciencia de hacer que las máquinas hagan cosas que requerirían de inteligencia si fueran hechas por hombres”. En el marco de las múltiples concreciones de esta idea general, desde una perspectiva “fuerte”<sup>18</sup> de la IA, algún sector defiende que, dada la correcta configuración de la relación entre *inputs* y *outputs*, una máquina tendría tanta inteligencia como los humanos (Copeland, 1993).<sup>19</sup> Esta es la visión más radical y, claro, también la más frecuentemente criticada.<sup>20</sup>

<sup>16</sup> Así, entre muchos, recientemente Grace (2018) se cuestiona sobre cuándo la inteligencia artificial va a exceder el desempeño humano.

<sup>17</sup> De reciente aparición, creciente desarrollo y prometedor futuro, tal como lo advertía ya hace unos años el filósofo sueco Bostrom (2014), en referencia a lo que llamó “superpoderes cognitivos”.

<sup>18</sup> Sobre la diferencia entre el sentido fuerte y débil de la inteligencia artificial, recientemente López de Mántaras (2019: 162) planteó que “una IA fuerte implicaría que un ordenador convenientemente diseñado no simula una mente sino que *es una mente* y por consiguiente debería ser capaz de tener una inteligencia igual o incluso superior a la humana. La IA débil, por otro lado, consistiría, según Searle, en construir programas que realicen tareas específicas y, obviamente sin necesidad de tener estados mentales”.

<sup>19</sup> En el mismo sentido, Pinker (2011: 11) considera que “el cerebro, nos guste o no, es una máquina”.

<sup>20</sup> Esta es la visión que inspira (¿y sigue inspirando?) el conocido debate entre la prueba de Turing y el cuarto chino de Searle, dos visiones antagónicas de la naturaleza de la inteligencia y, en general, de la mente humana. Son, por supuesto,

Quizás, hoy en día, el mayor reto para la IA fuerte resulte de la inexistencia de una generalizada o multipropósito y, por supuesto, el complejo asunto de la “conciencia reflexiva”.<sup>21</sup> Sin embargo, la discusión actual no parece ser si ello será posible —especialmente, la IA multipropósito—, sino cuándo.<sup>22</sup> Por eso, independientemente de la postura que se adopte en relación con este asunto desde la filosofía de la mente,<sup>23</sup> es menester estar al tanto de los pormenores de la investigación en este frente y, permanentemente, reflexionar sobre sus implicaciones para el derecho penal.<sup>24</sup>

En una cuestión relacionada —quizás, desde alguna perspectiva, la misma—, Sánchez del Campo (2020) también ofrece una reflexión actualizada acerca de la capacidad de agencia de las máquinas y la posibilidad de atribución de responsabilidad jurídica a los robots. Desde una creativa y provocativa postura, sugiere que:

En la medida en que los robots puedan tener una personalidad legal, también podrían cometer ilícitos civiles y/o penales y ser responsables de sus actuaciones. ¿Cuáles podrían ser las sanciones asociadas a los ilícitos criminales de los robots? Lo lógico sería que las penas se impusieran en función de la gravedad de tales actuaciones u omisiones, pudiendo categorizarse en distintos niveles dependiendo de si se trata de delitos menores o más graves, graduándose las sanciones en consecuencia (desde la suspensión temporal de las

---

demasiadas publicaciones al respecto como para, siquiera, intentar recoger una muestra significativa en esta nota. Sin embargo, conviene recordar que las posturas originales se encuentran, respectivamente, en Turing (1950) y Searle (1998).

<sup>21</sup> Según Kurzweil (2012: 27), “nosotros somos unas máquinas que se están conociendo a sí mismas o, en otras palabras, que tenemos una conciencia reflexiva”. Esto refleja toda una corriente de pensamiento que considera que este es el rasgo característico de la “inteligencia humana” y, me atrevería a decir, de la propia naturaleza humana. Dada la vastedad de la producción interdisciplinaria a este respecto, es imposible recogerla siquiera a modo de ejemplo en este limitado escrito, pero una buena referencia a la discusión —actualizada, además— podría encontrarse en Smith (2020).

<sup>22</sup> Recientemente, en relación con sus implicaciones para el derecho, Delgado de Molina Ruis (2020: 114) plantea que “Desde un enfoque técnico, hemos distinguido entre *artificial narrow intelligence* (ANI), *artificial general intelligence* (AGI) y *artificial superintelligence* (ASI). Todos los supuestos actuales de IA entran dentro del ámbito de la ANI, pero el reciente poder de aprendizaje de los algoritmos está inclinando la balanza a favor del paradigma AGI”.

<sup>23</sup> En algún momento concluí que una visión funcionalista de la mente era, probablemente, el camino que mayor capacidad de rendimiento tiene a la hora de ocuparse de los complejos asuntos que en la actualidad y, en un futuro, traerá la discusión acerca del libre albedrío. He hecho referencia aquí a algunos recientes (como el de las neurociencias) y actuales (como el de la inteligencia artificial). Pero me inclino a pensar que las bases filosóficas desde las cuales parte esta perspectiva permiten abordar un sinnúmero de otras cuestiones sobre la naturaleza de la mente humana, principalmente porque, como explica Block (1980: 175), “el funcionalismo responde a la cuestión metafísica sin responder a la cuestión ontológica”. Desde esta perspectiva, sea que la libertad se ejerza desde una realidad bioquímica, silico-eléctrica o, en un nivel mayor de abstracción, completamente inmaterial, debería seguir una serie de patrones comunes que permitieran esperar de quienes la ejercen un comportamiento característico. ¿Sería esto suficiente para el derecho penal? Habrá que someterlo ahora a prueba como hipótesis de trabajo y seguir el camino de la investigación para hallar respuesta.

<sup>24</sup> Tanto en el campo sustantivo como en el procesal. Recientemente, en este frente el profesor Nieva Fenoll (2018) ha publicado uno de los libros más conocidos en nuestra tradición continental sobre un asunto que, desde hace un buen tiempo, ha ocupado a la doctrina anglosajona: la relevancia de las neurociencias en la práctica del derecho.

actividades del robot hasta la reprogramación de su software o, más aún, la destrucción física del hardware y del software) (Sánchez del Campo, 2020: 398).

Aunque, acto seguido, quizás contrario a lo que uno esperaría, expresa que “los robots no pueden ser considerados como sujetos responsables de sus actos u omisiones, sino que lo será su fabricante o importador, si se considera que han funcionado defectuosamente o el propietario o usuario cuando aquellos sean meros instrumentos” (Sánchez del Campo, 2020: 398). Ello, por cuanto, según explica el autor, “una cosa es que el derecho reconozca personalidad y otra que atribuya capacidad de obrar” (Sánchez del Campo, 2020: 399).

De otra opinión es Hernández Giménez (2019), para quien la posibilidad de atribuir obligaciones se correlaciona con la capacidad de reconocer responsabilidad jurídica por su incumplimiento. En una contribución, también reciente, sobre este particular, pero desde la específica órbita del derecho penal, esta autora alega que el estado actual de desarrollo de la inteligencia artificial permite concluir que los robots adopten comportamientos que *ex ante* no son completamente previsibles por quien hace la programación inicial.<sup>25</sup> Aunque no lo plantea expresamente, en este contexto, la incertidumbre sobre el específico curso que adoptará el desarrollo autónomo de la máquina en virtud de sus capacidades de autoaprendizaje presenta un interesante punto de debate con respecto a la posibilidad de “regresar” la imputación a quien la fabricó.

Es, cuando menos, inspirador pensar que una institución ideada hace más de un siglo<sup>26</sup> por quien escribía en una época en la que la doctrina mayoritaria debatía sobre cuál era —representación o voluntad— el elemento psicológico que mejor caracterizaba a la culpabilidad,<sup>27</sup> sirva ahora para

<sup>25</sup> En sus palabras: “Hoy en día, el hecho de que los robots estén equipados con inteligencia artificial hace que sean capaces de aprender por sí mismos y de tomar decisiones de manera autónoma, cosa que necesariamente implica la posibilidad de que estos adopten un comportamiento imprevisible para el ser humano, por lo que no parece razonable culpar a su fabricante de los daños que se pudieran ocasionar como consecuencia de este comportamiento imprevisible. Dado que es el fabricante el que programa o ‘educa’ al robot, será responsable de los daños provocados por una mala programación (o educación) pero, sin embargo, su responsabilidad ya no será tal cuando los daños se hayan producido como consecuencia de la capacidad de aprendizaje de estos, ya que la autonomía del robot aumenta paralelamente al nivel de desarrollo de la inteligencia artificial. Asimismo, hay que tener en cuenta la posibilidad de que los robots puedan ser ciberatacados por criminales, cosa que los haría aún más imprevisibles para su fabricante”. (Hernández Giménez, 2019: 801-802)

<sup>26</sup> Según registra Reyes Alvarado (2005: 187), se trata de “un antiguo principio denominado ‘prohibición de regreso’, que había sido expuesto a comienzos del siglo XX en Alemania por Reinhart Frank como un mecanismo para impedir que la responsabilidad penal se extendiera de manera ilimitada (sobre los orígenes de la prohibición de regreso, *cf.* Herbert Tröndle, StGB, Vor § 13, Rd. 18b)”.

<sup>27</sup> Frank (2002: 25 y 27), desde hace más de un siglo, advertía que “Tampoco importa por el momento si la propia relación psíquica debe ser considerada como una voluntad o como una representación mental. Lo principal es la limitación del concepto de culpabilidad a la faz interna”, pues, según explicaba: “Sea como fuere, la doctrina dominante

elevant un punto de debate sobre la posibilidad de atribuir responsabilidad a agentes que obrarían, por completo, sin atributo psicológico alguno.

En definitiva, la cuestión acerca de la agencia (no ya, solamente, inteligencia) artificial encierra un debate de profundidades abismales, respecto del cual los penalistas no somos extraños, pero que, en muchos sentidos, nos desborda (aunque en alguna oportunidad histórica algún filósofo haya caracterizado famosamente a las personas a partir de su capacidad de “crear leyes”<sup>28</sup> [que no jurídicas, aclaro]). Sea como fuere, lo relevante de este asunto para los específicos fines de este proyecto es que esta cuestión, inevitablemente, motiva una discusión que, en mi opinión, es tanto o más interesante que aquella suscitada en la década pasada sobre los avances de las neurociencias.

De concretarse en las próximas décadas como probablemente ocurra un sistema autónomo de inteligencia artificial generalizada o, al menos, una prueba innegable de su proximidad, se abriría una nueva dimensión de la discusión. Y es, verdaderamente, una realidad emocionante el debate, más allá de las neurociencias, sobre libertad artificial a la cual, como siempre, el derecho deberá hacer frente para actualizarse y, quizás, hasta reinventarse.<sup>29</sup>

---

encuentra en la esencia de la culpabilidad una relación psíquica del autor con algo que está ubicado fuera de su personalidad”. Para este autor, la culpabilidad no expresaba —al menos, no enteramente— un fenómeno psicológico interno del individuo, sino que comprendía una serie de circunstancias que nada tienen que ver con ello (las llama “concomitantes”). En contra de lo que consideraba la doctrina mayoritaria de su época, propuso el caso de la exclusión de culpabilidad en los casos de estado de necesidad para demostrar que esta no se reducía al fenómeno subjetivo que coincidía con la imputación dolosa, algo que eventualmente llevaría a la doctrina a rechazar esta concepción psicológica de la culpabilidad. Así lo reconoce, expresamente, Mir Puig (2016: 546). El rechazo de la culpabilidad en sentido psicológico, mayoritaria para su época, llevó a que la doctrina contemporánea lo reconozca como uno de los principales promotores de la normativización de la culpabilidad. Según Mir Puig (2016: 620), “La teoría normativa de la culpabilidad así lo hizo, al acudir a la idea de reprochabilidad. Ésta se hizo depender en un principio —cuando todavía se defiende la sistemática causalista— del dolo y la imprudencia, pero también de otras circunstancias que pueden llegar a excluir la reprochabilidad del hecho doloso o imprudente. Frank, uno de los creadores de la teoría, habla en este sentido de ‘circunstancias concomitantes’”.

<sup>28</sup> Dennett (1998: 145) cita lo expuesto por Locke: “Persona es un término forense, apropiación de acciones y de sus méritos; y además solo pertenece a agentes inteligentes, capaces de crear leyes, felicidad y tristeza. Esta personalidad extendida asimismo más allá de la presente existencia de lo que es pasado solo por su consciencia, por lo cual se convierte en preocupada y explicable”.

<sup>29</sup> Ya algunas facultades han emprendido un papel protagónico en este frente. Harvard, por ejemplo, ofrece en su curso “Ciencia informática para abogados” (CS50 for Lawyers) un módulo destinado a los “Retos en la intersección entre Derecho y tecnología” (Challenges at the Intersection of Law and Technology), en el cual expresamente se abordan interesantes problemas como las implicaciones de la agencia virtual en espacios de interconexión y experiencia común de cara a los deberes positivos y negativos en el relacionamiento interpersonal, o los desafíos que supone el futuro de la impresión 3D para la política criminal. A propósito, Thierer y Marcus (2016).

#### IV. PERSONALIZACIÓN DE INTERNET, “CÁMARAS DE ECO” Y ACCESO NORMAL A LA PROHIBICIÓN: DESAFÍOS Y PERSPECTIVA A FUTURO

Hace casi un siglo y medio, Beling (2002: 114) alegaba que toda la culpabilidad y ambas de sus especies encerraban, en esencia, un reproche al sujeto inspirado en una “interna actitud con respecto a la antijuridicidad de su conducta”,<sup>30</sup> que se presenta “cuando sabe que, conforme a nuestro orden jurídico, no debe obrar como obra, es decir, que se halla en contradicción con la voluntad común dominante en el Estado”.

Medio siglo después, desde un planteamiento sustancialmente diferente, Welzel se ocupó, con más detalle,<sup>31</sup> de distinguir entre el error de tipo y el de prohibición a partir del objeto del conocimiento en cada caso. En particular, sobre el (des)conocimiento del fundamento de la norma de prohibición, este autor explicaba:

El error sobre la validez de una norma del derecho penal (error de validez, a diferencia del error sobre el contenido de la norma, de que hemos tratado hasta ahora) es una variedad del error de prohibición. El autor conoce la prohibición, pero cree que no es válida, porque

<sup>30</sup> Beling (2002: 105) afirmaba que “lo que se reprocha al autor es su interna actitud con respecto a la antijuridicidad de su conducta. Conforme a ello, cada una de ambas graduaciones de la culpabilidad es pues también unitaria: el *dolus* significa reprochar al autor el hecho de no haberse detenido ante el pensamiento de estar obrando antijurídicamente; la culpa, reprochar al autor el hecho de desconocer la antijuridicidad de su conducta, debiendo no haberla desconocido”. Para otros autores, en cambio, la verificación del conocimiento es instrumental; me explico: es requerido solo porque, sin este, no puede predicarse una “resolución de voluntad antijurídica”. Para Welzel (2006: 161 y 174), “nunca se le puede reprochar al autor la resolución de voluntad antijurídica si no podía tener conciencia de la antijuridicidad. La doctrina se ha esforzado, por ello, desde hace mucho tiempo, por imponer también el principio de la culpabilidad en el error de prohibición”; y, más adelante, expone que el “Objeto del juicio de reproche de la culpabilidad es la resolución de voluntad antijurídica: ésta le es reprochada al autor en la medida en que podía tener conciencia de la antijuridicidad de su acción y en que dicha conciencia podía convertirse en un contramotivo determinante de sentido”.

<sup>31</sup> Probablemente, porque en el “Nuevo Sistema” las consecuencias de ambos errores son radicalmente distintas, cosa que no ocurría —al menos, no en la misma medida— en el esquema clásico causalista, en el que ambas clases de error se referían a una misma categoría del delito y, de hecho, a una misma institución (el dolo). De forma expresa, Beling (2002: 111-112) expuso que “el concepto de dolo requiere que el autor tuviera la conciencia de la antijuridicidad de su acción (...) Su justicia se deduce de que al que actúa en bona fide, al que cree no contrariar el orden jurídico, no se le puede reprochar nada o sólo puede reprochársele su inadvertencia, lo cual no corresponde al cuadro del dolo, como grado máximo de la culpabilidad. La intención sólo tiene sentido dirigiéndose al tipo de ilícito y no solo a lo externamente típico, con prescindencia del contenido ilícito”. Sobre la explicación histórica del interés en distinguir adecuadamente los contornos de uno y otro tipo de error y, en particular, sobre la evolución y fundamentos del tratamiento del error de prohibición (recurriendo, incluso, a explicaciones socioeconómicas), vale la pena revisar lo desarrollado por Jakobs (1999: 1062); en particular, es llamativa la referencia a la sociedad burguesa como un factor histórico en esta cuestión: “La fundamentación debe ser otra: la obligación de cada uno de procurarse él mismo conocimiento del derecho era el precio que había que pagar por la libertad de movimientos en la sociedad burguesa. Especialmente al abandonar las modalidades tradicionales de actividad económica y generarse con ello una necesidad de seguridad a la hora de mantener contactos bastante anónimos, el riesgo de que los demás errasen acerca de sus obligaciones jurídico-penales era intolerable; por ello, quien erraba debía soportar el riesgo”.

según su opinión infringe una norma jurídica positiva de rango superior, por ejemplo, de la Constitución, o un principio metapositivo. Aquí rigen también las reglas del error de prohibición. Si el error es disculpable, el hecho queda impune; si no es disculpable, el hecho sigue siendo punible (Welzel, 2006: 185).

Ambas posturas, pese a partir de esquemas en muchos sentidos radicalmente opuestos, coinciden en la relevancia del conocimiento que el agente tiene, no solamente sobre la existencia de la prohibición positiva, sino sobre su fundamento. Beling se refería a una “voluntad común dominante en el Estado”, mientras que Welzel a “un principio metapositivo” (¿como el de la soberanía de la voluntad popular?), pero ambos coincidían en que hay algo *más allá de la norma positiva* cuyo conocimiento resulta relevante para determinar el reproche a la persona por el injusto típico.

Esto es muy interesante, pues nos hace preguntarnos: ¿qué consecuencia tendría, si alguna, que alguien, fruto de una visión sesgada por la personalización de su experiencia en internet, estuviera genuinamente convencido de que todos (o la abrumadora mayoría de las personas) opinan, como él, que la norma de prohibición es ilegítima? ¿Qué pasaría si, fruto de la interacción limitada con personas que comparten sus mismos intereses, resultara profundamente convencido de que estos representan a la voluntad mayoritaria e, inspirado en su soberanía, se tomara violentamente el orden público en protesta? ¿Podría decirse que, en esas condiciones, es consciente de la antijuridicidad de su conducta o que tiene, siquiera, un acceso normal a la prohibición?

Creo que es fundamentalmente acertada la distinción que hace Roxin (1997) entre la delincuencia por convicción<sup>32</sup> y el error sobre la validez de la norma.<sup>33</sup> Ciertamente, no es el mismo caso el del “foráneo” que el de quien yerra, precisamente, sobre esta condición. Me explico.

<sup>32</sup> El autor plantea: “Si se contempla como objeto de la conciencia de la antijuridicidad la prohibición o el mandato jurídicos, de ello se deduce que el sujeto que actúa por motivos de conciencia o por convicción no puede invocar el §17, en la medida en que sabe simplemente que el ordenamiento jurídico prohíbe tal conducta. Por tanto, quien comete un allanamiento de morada (§123) o unas coacciones (§240), porque está convencido de que debe manifestarse de este modo contra abusos sociales o políticos, no actúa en error de prohibición si sabe que el ordenamiento jurídico desaprueba el hecho que él considera necesario. (...) Con ello no se dice que la voz de la conciencia o incluso simplemente las convicciones respetables sean irrelevantes para la valoración jurídica-penal. Pueden tener consecuencias a favor del sujeto en la medición de la pena e incluso dar lugar a una causa de exclusión de la culpabilidad en casos concretos (al respecto §22, nm. 122 ss.). Pero eso es harina de otro costal”. (Roxin, 1997: 868-867)

<sup>33</sup> Al respecto, Roxin (1997: 873) menciona que “Una forma menos frecuente de error de prohibición es el error de validez. En él el sujeto conoce la norma prohibitiva y por regla general incluso la norma penal, pero la considera nula, p. ej., porque al órgano que la dictó le faltaba la competencia legislativa, o porque el precepto atenta contra un derecho fundamental o contra el principio de determinación. Se ha de juzgar este caso como error de prohibición en tanto en cuanto quien actúa invoque, como en los ejemplos citados, causas de nulidad que también estén reconocidas por el ordenamiento jurídico. Otra cosa sucede cuando el sujeto considera válida una ley conforme a los parámetros

Uno de los planteamientos más conocidos sobre el tratamiento de los “foráneos” es el de Jakobs.<sup>34</sup> En breve, alega que su culpabilidad no es, en esencia, distinta de la de la delincuencia ordinaria, sino que deriva también de la comunicación de una (extrema) infidelidad con el ordenamiento.<sup>35</sup> En ese contexto, concluye que “no es necesario exigir más que la cognoscibilidad de un comportamiento que atenta contra el ordenamiento dado (o en el caso de los motivos bajos: que atenta de forma cualificada), en especial, la no cognoscibilidad de la preferencia de ese ordenamiento” (Jakobs, 2008: 133).

En el caso ahora planteado, creo que el problema no es, exactamente, el mismo: el foráneo se reconoce a sí mismo como foráneo en relación con aquella sociedad que repudia; quien, en cambio, ha llegado al convencimiento de que la abrumadora mayoría rechaza la legitimidad de una norma positiva de prohibición, no se representa como foráneo (de hecho, llegaría a decir que es el más *intraeus* de todos). Por eso, creo que, en relación con el problema planteado, vienen al caso consideraciones adicionales.

Dejando de lado, por ahora,<sup>36</sup> las cuestiones relativas al objeto del conocimiento relevante para la culpabilidad discusión que, como se ha visto, dista mucho de ser novel, en lo que sigue haré referencia únicamente a un asunto específico que atañe a la *forma* en que se interactúa con dicho objeto. A continuación, me referiré al problema de los sesgos cognitivos asociados a la personalización de internet que interfieren en el reconocimiento del fundamento de la norma de prohibición, y el problema de las “cámaras de eco” para el derecho penal.

Tradicionalmente, los modelos de imputación del injusto han partido de la idealización del hombre medio, a veces discriminado según la posición del autor. Así, por ejemplo, a lo largo del siglo pasado, la comprobación de

---

del ordenamiento jurídico, pero no la estima vinculante para sí por razones políticas, ideológicas, religiosas o de conciencia. Las razones de esa índole no modifican en nada la conciencia del carácter prohibido de la propia acción, que es lo único decisivo”.

<sup>34</sup> Una línea que ha construido a través de varias publicaciones durante más de tres décadas, muchas de ellas traducidas al castellano.

<sup>35</sup> Con detalle, Jakobs (2008: 128) señala: “En la determinación de los foráneos de lo que se trata es de saber si el autor considera que un ordenamiento, más o menos ajustado a las exigencias de la modernidad, es la estructura de una banda de ladrones, es decir, es un ordenamiento sin justicia, para cuya represión a él le debe parecer permitido todo lo que tiene una fuerza simbólica (y en esa medida también tiene un efecto concreto). El mencionado ‘debe’ es, sin embargo, un ‘debe’ concreto psicológico: al autor nada le parece un motivo para acercarse al Derecho. Esta es la infidelidad al derecho mayor posible y, por esto, es culpabilidad, en la medida en que el conflicto no se pueda solucionar de otra forma. En tratándose de un suceso aislado que no afectará más el ordenamiento (...), la condición síquica puede tratarse como una deficiencia mental, precisamente, social”.

<sup>36</sup> Con seguridad, será un asunto en el que deberé profundizar en el desarrollo de mi investigación.

la violación objetiva del deber de cuidado estuvo dominada en general por la aparición de modelos como el *hombre prudente*,<sup>37</sup> el *hombre diligente* y el *hombre inteligente*;<sup>38</sup> o, en la misma línea, el *buen padre de familia*, el *buen médico* y el *buen hombre de negocios*, entre varios otros modelos específicamente asociados al rol ocupado por el agente.<sup>39</sup> Aunque cada modelo ofrecía algunas particularidades propias del rol respectivo, todos tenían en común que derivaban de un intento por determinar el grado de concordancia o discordancia de la conducta del agente con el comportamiento debido y como una necesidad de responder la siguiente pregunta: “Idealmente, ¿cómo se habría comportado el hombre promedio, puesto en la posición del autor?”<sup>40</sup>

Sin embargo, todos estos modelos tienen en común la aceptación de una premisa tácita que fundamenta todo el juicio de concordancia o discordancia anteriormente referido: que *el hombre, idealmente, se comporta de manera racional*. Este postulado, propio de la economía clásica, que concibe al hombre como *homo œconomicus* (Mac Adams, 2008, como se citó en Silva Sánchez, 2016), no es aceptado universalmente y, de hecho, la perspectiva conductual del derecho y la economía —*Behavioral Law and Economics*— parte de un postulado distinto. Según este sector de la doctrina contemporánea, el comportamiento del ser humano no está determinado en todos los contextos por la estricta reflexión racional. Por el contrario, existen contextos específicos en los cuales la interacción del individuo con dinámicas preestablecidas altera su juicio y condiciona su proceder voluntario, bien por un influjo en la razón (*bounded rationality*), en la voluntad (*bounded willpower*) o en el autointerés (*bounded self-interest*) (Silva Sánchez, 2016: 248-249).

Los influjos del contexto sobre el sujeto, mencionados anteriormente, producen los llamados “sesgos cognitivos”,<sup>41</sup> que pueden conllevar una alteración sistemática en la percepción de la realidad que impide al

<sup>37</sup> Sobre el importante consenso que se logró consolidar sobre esta figura en el siglo XX, Maurach (1962) y Jescheck (1993), quienes dan fe de la predominancia de tal figura durante la segunda mitad del siglo pasado en la doctrina alemana. Igualmente puede verse la STS 18-10-1979 (A. 3737) y 16-6-1987 (A. 4955), que caracterizan una corriente sumamente decantada en el Tribunal Supremo español. Entre muchos otros que han propugnado en algún momento por esta figura, se encuentran: Welzel (1997), Kaufmann (2006), Jescheck (1993), Roxin (1997) y Muñoz Conde (2010).

<sup>38</sup> Para un amplio panorama de las diversas denominaciones que ha tenido esta figura a lo largo de la historia del derecho penal, véase Reyes (2005). Por su parte, Silva Sánchez y Varela (2013: 271) se refieren a “persona razonable como expresión del hombre medio”.

<sup>39</sup> Con mayor detalle, Díaz Arana (2014).

<sup>40</sup> En contraposición al modelo de previsibilidad individual en la determinación del deber de cuidado, que toma como punto de partida la verificación de todas las posibilidades de evitación del resultado lesivo con las que contaba el agente en concreto. Al respecto, Díaz Arana (2017).

<sup>41</sup> Silva Sánchez y Varela (2013: 272) señala: “Según se ha indicado, los sesgos cognitivos se generan en el razonamiento intuitivo, sin demasiada reflexión. Parece, pues, que bajo su influencia el sujeto obra, si bien de forma voluntaria, de un modo automatizado sobre la base de una serie de errores sistematizados por la habituación”.

agente formarse una imagen acertada de esta<sup>42</sup> y, en consecuencia, ordenar su comportamiento.<sup>43</sup> En el ámbito de la empresa, según explica Silva Sánchez (2016), los sesgos cognitivos propios de las dinámicas de grupo suelen aparecer, entre otras, a través de variadas formas, como el exceso de confianza,<sup>44</sup> de optimismo, de ilusión de control, de confirmación, de conformidad y de obediencia a la autoridad.

Pero, también, existe otro notorio espacio de interacción altamente organizada en la que este fenómeno se presenta. En particular, me refiero a la interacción virtual desde la óptica del problema de la personalización de internet.

Según reportó el año pasado el *New York Times* (2019):

[E]n un país llevado al borde por la crisis económica y política, los algoritmos de YouTube pudieron haber jugado un rol decisivo en el auge de Bolsonaro. La función de recomendación del sitio impulsa los videos marginales a la corriente principal y puede ayudar involuntariamente a difundir conspiraciones y desinformación sobre enfermedades peligrosas, poniendo en peligro la salud pública. (Jordan y Shorland, 2019).

En efecto, los investigadores registraron que el problema relacionado con los algoritmos de maximización de la experiencia del usuario de YouTube, en el contexto electoral de Brasil, radicalizó a la población y desbordó el campo de la intervención en política, llegando a incidir determinantemente en la mayoría de los ámbitos de la vida cotidiana, como la educación o la salud pública.<sup>45</sup>

<sup>42</sup> Parker (1993: 741), quien, de manera provocativa, abre su brillante disertación reconociendo para la tradición anglosajona un punto al que aquí también nos hemos referido ahora: el carácter enigmático de la culpabilidad (en la tradición anglosajona, *mens rea*). Textualmente: “The criminal law doctrine of *mens rea* the guilty mind has been something of an enigma to the economic analysis, despite its important place in legal thinking”.

<sup>43</sup> Saavedra Velasco (2011: 64) afirma que “se entenderá por sesgo a aquella circunstancia por la cual la ocurrencia de un error no resulta verdaderamente un evento aleatorio, sino que este se presenta en la realidad de manera sistemática”. También De la Rosa Rodríguez y Sandoval Navarro (2016). Conviene señalar, en este momento, que por sesgos se entienden aquellas reglas cognitivas que, inconscientemente, todo ser humano aplica al procesar la información que recibe del exterior, y que permiten “reducir las tareas complejas de asignar probabilidad y predecir valores a operaciones de juicio más simples” (De la Rosa Rodríguez y Sandoval Navarro, 2016: 148). De igual forma, son llamados procedimientos heurísticos por tratarse de procedimientos mentales que todos los seres humanos llevamos a cabo cuando percibimos y analizamos una determinada situación.

<sup>44</sup> Según Cortada de Kohan (2008: 71-72), “El sesgo de sobreconfianza puede ser definido como un error sistemático de calibración subjetiva de éxito en la toma de decisiones bajo incertidumbre (Camerer y Lovoal, 2000; Oskamp, 1965). Este error consiste en la sobreestimación del éxito de las propias decisiones”.

<sup>45</sup> Tras su investigación, Fisher y Taub (2019) reportaron que, para el caso de las elecciones de Brasil, “una y otra vez, los videos promovidos por el sitio han volcado elementos centrales de la vida cotidiana. Los maestros describen aulas que los estudiantes que citan en los videos de conspiración de YouTube (...). Algunos padres miran a ‘Dr. YouTube’ para obtener consejos de salud, pero en su lugar reciben información errónea peligrosa, lo que obstaculiza los esfuerzos de la nación para combatir enfermedades como el zika. Los videos virales han provocado amenazas de muerte

Esto, en el caso de Brasil, pero tanto más se puede decir de Inglaterra,<sup>46</sup> Estados Unidos,<sup>47</sup> España<sup>48</sup> o Colombia,<sup>49</sup> entre muchos otros países.

Sin negar lo anterior, más allá del inagotable registro del asunto en medios, para los fines de este escrito conviene centrar los esfuerzos en la revisión del estado de la cuestión en la bibliografía autorizada. A este respecto, aunque es forzoso manifestar, de entrada, que este asunto es un terreno inexplorado en el campo específico del derecho penal, en una revisión interdisciplinaria más amplia sí se encuentran aportes interesantes.

En una reconocida publicación de 2016, investigadores de IUSS Pavía (Italia), CSSLab IMT Lucca (Italia), ISC CNR Roma (Italia) y NICO NW Evanston (EE. UU.) analizaron, desde una perspectiva conceptual y estadística, la “polarización de usuarios en Facebook y YouTube” (Bessi, 2014: 259-268). Desde esta aproximación al fenómeno referido como “cámara de eco”, esto es, la tendencia natural de los usuarios de internet a seleccionar la información que sustenta y adhiere a sus creencias y formar grupos polarizados con gente que comparte su misma perspectiva (Brown *et al.*, 2007, como se citó en Bessi, 2016), los investigadores afirman que:

[E]l sesgo de confirmación juega un papel esencial en la difusión de rumores en línea. Sin embargo, en las redes sociales, diferentes algoritmos favorecen contenido personalizado de acuerdo al gusto de cada usuario —me explico: muestran al usuario puntos de vista con los que ellos ya están de acuerdo. (...) [Esto] lleva al surgimiento de comunidades polarizadas en las cuales los usuarios adquieren información confirmatoria e ignoran contenido en contrario. (Bessi, 2016: 3)

Tras la recopilación de la información estadística de 12 millones de usuarios y su presentación gráfica, contrastándola entre sí y analizando sus puntos de coincidencia en cada caso, los investigadores reportaron interesantes hallazgos. De entrada, encontraron que “la agregación de usuarios en torno a narrativas opuestas lleva al surgimiento de cámaras de eco. Una vez dentro de tales comunidades homogéneas y polarizadas, los usuarios que apoyan las narrativas se comportan de la misma manera” (Bessi, 2016: 4). Pero, además, en relación con la “predictibilidad de la polarización de los

---

contra defensores de la salud pública. Y en política, una ola de estrellas derechistas de YouTube se postuló para el cargo junto con Bolsonaro, algunos ganando por márgenes históricos. La mayoría todavía usa la plataforma, que gobierna la cuarta democracia más grande del mundo a través de los ‘trolls’ y la provocación perfeccionadas en Internet”.

<sup>46</sup> Moore (2019) desarrolla el tema respecto de Boris Johnson.

<sup>47</sup> Recientemente, a propósito del *impeachment* (Stelter, 2020).

<sup>48</sup> Borreguero (2018) habla sobre la desinformación en España.

<sup>49</sup> En Colombia, Garavito (2016) desarrolla el tema en términos de paz y las cámaras de eco.

usuarios”, hallaron evidencia suficiente para sostener que “la interacción temprana de usuarios con contenidos es un predictor acertado del apego y preferencia hacia una comunidad y, por lo tanto, a las cámaras de eco” (Bessi, 2016: 7). Finalmente, el grupo de investigadores coincidió en concluir que “[t]al comportamiento común nos permite derivar un modelo de aprendizaje estadístico para predecir con una buena precisión si un usuario se polarizará hacia cierta narrativa o si variará entre contenidos apoyando diferentes narrativas” (Bessi, 2016: 11).

Otra notable contribución, más reciente, a la cuestión que ahora nos ocupa, viene de la mano de una alianza entre el equipo de informática y el de comunicaciones de la Universidad de California (EE. UU.) (Hilbert *et al.*, 2018: 260-275). Según reportaron, encontraron evidencia que permite concluir que el algoritmo de selección de videos de YouTube está conectado a emociones humanas a través de los términos empleados como criterio de búsqueda.

Más específicamente, a través de herramientas de aprendizaje automatizado<sup>50</sup> profundo (*deep learning*), basado en la tecnología de procesamiento natural del lenguaje adquirida por IBM, el reconocido proyecto Watson,<sup>51</sup> los investigadores identificaron los videos que apelaban a una o varias emociones en particular, así como la reacción que su visualización producía en los usuarios, y apreciaron cómo el algoritmo se alimentaba a sí mismo a través del reconocimiento de lenguaje emocional empleado por el usuario en sus búsquedas, para presentarle videos con una específica carga emocional.<sup>52</sup>

Al referirse a “las emociones de la polarización”, muestran cómo el estado emocional del usuario es identificado por el algoritmo de recomendación y, a su vez, el usuario ve alterado su estado emocional como resultado de seguir las recomendaciones del algoritmo.

Esto que llaman “transferencia entrópica de emociones” del algoritmo respecto del humano, al detectar sus emociones, y del humano respecto del algoritmo, al alterar su estado emocional por efecto de sus recomendaciones

<sup>50</sup> Sobre la utilización de la inteligencia artificial en la teoría de la información, véase Mackay (2003).

<sup>51</sup> Específicamente para Watson Developer Cloud: <https://www.ibm.com/watson/developer>

<sup>52</sup> Hilbert *et al.* (2018: 11) indican que “We used *AlchemyLanguage* from the IBM Watson Developer Cloud (now called ‘*Watson Natural Language Understanding*’) to execute a sentiment analysis that evaluated the feelings attached to the videos based on the scraped title, description, and transcript. *AlchemyLanguage* is a collection of APIs that offer text analysis through natural language processing. Before it was acquired by IBM in 2015, it was known as *AlchemyAPI*, a deep learning *machine learning* tool for natural language processing (specifically, semantic text analysis, including sentiment analysis). It evaluates positiveness and negativness (valence) on a scale from -1 to +1 and assigns values between 0 and 1 to the presence of anger, fear, disgust, joy, sadness (both to the third digit)”.

es una idea muy poderosa: en algún grado de abstracción, sugiere una especial relación entre dos *res cogitans*, *software* y espíritu. Esto, por sí solo, debería ser suficientemente llamativo para cualquier filósofo de la mente.

Pero, dejando esas cuestiones de lado por ahora (otra vez), quizás lo más interesante a efecto de lo que ahora nos ocupa sea mostrar la incidencia de esta realidad en la polarización y generación de cámaras de eco. Los investigadores tomaron como base de experimentación las elecciones presidenciales de EE. UU. en 2016 (Clinton vs. Trump) y llegaron a algunas conclusiones que bien pueden extrapolarse. En particular, es destacable la forma novedosa de aproximarse a la cuestión de la personalización de internet desde la transferencia emocional.<sup>53</sup> Resulta especialmente meritoria la reflexión final que ofrecen en torno a la importancia de comprender esta dimensión para entender el alcance real de los algoritmos de personalización y, en ese mismo camino, orientarlos de una forma socialmente responsable. En sus palabras (Hilbert *et al.*, 2018: 18): “Uno de nuestros hallazgos fue que el contenido cargado con emociones alegres conduce a la polarización emocional, mientras que la tristeza y el miedo producen convergencia. Esta es una idea útil en la búsqueda del diseño de algoritmos de recomendación socialmente más responsables que apunten a mitigar los efectos de la polarización de las burbujas de filtro actuales”.

Todavía más reciente es la publicación en *Nature* de un estudio de 1.2 millones de usuarios de Facebook que se ocupa, expresamente, de analizar el impacto de los sesgos de confirmación y de conformidad atribuibles a las cámaras de eco en los procesos de formación (y abandono) de opiniones en usuarios con narrativas conflictivas (Brugnoli, 2019). En su conclusión, fueron más allá que los anteriores estudios en el sentido de identificar subgrupos dentro de cámaras de eco y se reportó lo siguiente:

[D]emostramos cómo durante las interacciones sociales la fuerza del sesgo de confirmación es más fuerte de lo que uno podría pensar (...) fomentando la formación de subgrupos altamente polarizados dentro de la misma cámara de eco. El hecho de que los usuarios polarizados tienden a permanecer confinados dentro de grupos de muy pocas páginas, incluso cuando los vecindarios correspondientes están activos en varias fuentes de noticias, sugiere que el mecanismo de búsqueda de refuerzo limita la influencia de los vecinos y, principalmente, impulsa la selección y la difusión de contenidos incluso dentro de grupos de personas de ideas afines.  
(...)

<sup>53</sup> Según explican, “puede ser que la bien establecida distinción entre influencias individuales y sociales en la polarización en línea (Baksy *et al.*, 2015) no sea la forma correcta de escapar la cámara de eco” (Hilbert *et al.*, 2018: 18).

Finalmente, investigamos los efectos de la acción conjunta del sesgo de confirmación y la influencia de los compañeros (...). Nuestros hallazgos revelaron que los usuarios polarizados refuerzan sus creencias preexistentes al aprovechar la actividad de sus vecinos de ideas afines, y esta tendencia crece con la participación del usuario, lo que sugiere que la influencia de los compañeros actúa como un apoyo para la búsqueda de refuerzo (Brugnoli *et al.*, 2019: 15).

En estos momentos no cuento con la información suficiente para afirmar ni descartar las conclusiones arriba expuestas sobre predictibilidad del comportamiento humano, transferencia emocional o el influjo de los sesgos de confirmación y conformidad en la segregación de subgrupos al interior de una misma cámara de eco, ni para definir con claridad el alcance que llegarán a tener estas cuestiones en la dogmática penal. Sin embargo, sí puedo hacer notar dos cosas. La primera, que la “homofilia”, esto es, “la tendencia de usuarios de agregarse en torno a intereses comunes” (Anagnostopoulos *et al.*, 2015: 355-356), es un fenómeno que ha sido ampliamente discutido en el contexto de la personalización de internet<sup>54</sup> durante ya varios años,<sup>55</sup> desde distintas dimensiones,<sup>56</sup> en el marco de un debate plenamente vigente.<sup>57</sup> Y la segunda, que el hecho de que ya haya sido planteada la cuestión (a mi juicio, seriamente), anticipa que es inevitable que, a la larga, llegue a ser objeto de debate en el derecho penal.<sup>58</sup>

Si a estas “novedades” se les carga con el viejo debate sobre la relevancia del conocimiento sobre el fundamento de la prohibición en la determinación del reproche por el injusto, estoy seguro de que propuestas interesantes podrán alcanzarse en torno a problemas que, con toda seguridad, nos acompañarán en el futuro.<sup>59</sup>

<sup>54</sup> Este punto en particular ha sido abordado por los siguientes autores: Bessi (2015), Quattrociocchi (2019) y Zollo (2016).

<sup>55</sup> Hace casi una década, Aiello (2012) y otros ya mostraban que, a través de un análisis puramente basado en el agrupamiento (“clustering”) de intereses comunes como tecnología de aprendizaje automatizado, se podrían “predecir” con alto grado de fiabilidad las amistades en redes sociales. Ahora bien, ya en lo que se refiere a las bases teóricas de lo que hoy inspira el debate, debe reconocerse que casi coincidente con el nacimiento del término “inteligencia artificial” en la famosa conferencia de Dartmouth en 1956 se pueden trazar antecedentes que remontan hasta mediados del siglo pasado, como lo expone Attneave (1959).

<sup>56</sup> Desde sus dimensiones colectivas en política, en donde mayor producción específica ha habido, destaca Garret (2013) o, en general, en la sociedad Bessi (2016), hasta su dimensión individual (Zollo, 2015).

<sup>57</sup> Algunas de las principales contribuciones tienen escasos meses, como es el caso de Choi (2020), Baldwin-Philippi (2019), Cota (2019) y Ortiz (2019).

<sup>58</sup> Piénsese que el famoso experimento de Libet fue realizado en 1983 y publicado en 1985, pero, en nuestra tradición continental, la discusión jurídico-penal respectiva apenas inició después de 2000.

<sup>59</sup> Sobre las cámaras de eco, por ejemplo, Anagnostopoulos *et al.* (2015: 2) acertadamente destacan su carácter sistémico: “[t]ales grupos son muy similares: presentan un fuerte homofilia (sus usuarios tienden a interactuar con usuarios con una polarización similar y a consumir información con similares patrones). Además, la estructura interna de la red social es estadísticamente similar”. Abordar estos espacios desde la identificación de sus coincidencias independientemente de cuál sea la narrativa que adopten sobre algún asunto específico: elecciones, independencia, política

Tal es el caso, por ejemplo, de la manipulación sistemática de los sentimientos de los votantes en un certamen democrático, de la capitalización de las cámaras de eco para radicalizar opiniones en asuntos de política pública, del recurso (aparentemente incontenible) a la desinformación masiva<sup>60</sup> para promover el desorden generalizado o la oposición a alguna política pública,<sup>61</sup> y tantos otros que, me temo, no desaparecerán en el corto plazo.

En cualquiera de estos contextos, alguna persona puede llegar a encontrarse inadvertidamente en una cámara de eco que determine las condiciones de acceso a la información e, influenciado por los sesgos correspondientes, errar de modo personalmente invencible sobre el fundamento de la norma de prohibición. Este caso no es de un delincuente por convencimiento como el foráneo sino, justamente, el de un convencido por el error.

## V. REFLEXIÓN FINAL

En las líneas precedentes, he agotado mi esfuerzo por recapitular el estado de la discusión en torno al libre albedrío en la culpabilidad jurídico-penal, incluyendo —pero no limitándose a— la cuestión sobre las neurociencias. He presentado mis consideraciones desde distintas perspectivas el pasado reciente, presente y probable futuro que abordan problemas tradicionales desde enfoques contemporáneos. Con ello, he querido transmitir la relevancia de la cuestión para el derecho penal y por qué considero que sigue existiendo un importante espacio en el que pueden hacerse contribuciones significativas al estado actual y venidero de la discusión.

Este tema fue muy importante. Desde una dimensión histórica, sirvió para recordar que el debate sobre el libre albedrío no parece ser uno de

---

económica, etc., permite apreciar claramente sus dinámicas fundamentales y, desde ahí, diseñar políticas públicas para contrarrestar sus efectos indeseados.

<sup>60</sup> Específicamente, sobre el penoso fenómeno de las *fake news*, que genera por sí solo tantos retos para la política criminal y, en general, política de seguridad pública, resulta interesante la aproximación de Anagnostopoulos *et al.* (2015), para quienes “homofilia y polarización podrían ser las métricas clave para identificar las comunidades de una red social donde los rumores falsos o engañosos son más probable que se extiendan”. Este tipo de consideraciones, de carácter técnico, deben orientar a las autoridades en la difícil lucha contra este peligroso enemigo. Ello, sin desconocer que la solución definitiva deberá apoyarse, también, en algún ámbito de autorregulación (como el que ya hemos empezado a ver en Twitter o YouTube, al promocionar contenido de autoridades en sus algoritmos de recomendación).

<sup>61</sup> Una de las contribuciones más recientes se refiere (cómo no) al debate en la opinión pública sobre política pública de vacunación a propósito del coronavirus (Cossard, 2020).

aquellos que se “superen”. Ahora, avances en inteligencia y agencia artificial nos enfrentan a un panorama en el que los descubrimientos sobre la operación del cerebro bioquímico podrían llegar a ser triviales para la atribución de responsabilidad. De otro lado, ya hay evidencia de que la propia configuración de la interacción en línea fomenta la implantación en las personas de sesgos que, aunque no necesariamente tienen una base biológica, sí pueden llegar a condicionar los procesos individuales de toma de decisiones e interferir con las posibilidades de acceder normalmente a la prohibición.

Como se ve, pese a la importancia que tiene la disputa sobre las neurociencias, hay muchas otras cuestiones sobre culpabilidad jurídico-penal que escapan a su alcance. De hecho, son de tal relevancia que deberían mostrar que aquella se refiere apenas a una de varias dimensiones del problema. Las últimas dos décadas impulsaron valiosos avances en el pensamiento dogmático, pero aún quedan importantes retos en una historia que está lejos de concluir. Los avances más recientes, nuevamente, obligan a repensar algunas categorías y a actualizar conceptos determinados.

Pero esto es natural: en la historia de la culpabilidad siempre han quedado plasmadas las vicisitudes de la evolución del pensamiento en torno a la naturaleza del hombre y su vida en sociedad. Por eso, es muy inspirador leerla y comprender, en paralelo, que ella cuenta la historia del propio ser humano. De ahí que convenga repasar constantemente los avances más recientes en esta historia y —¿por qué no?— dejarse llevar por algunas conjeturas sobre los próximos capítulos. Parece que lo mejor aún está por venir.

---

## VI. FUENTES DE CONSULTA

- Aiello, L et al. (2012). “Friendship prediction and homophily in social media”. *ACM Transactions on the Web (TWEB)*, 6, 2, 5-39.
- Alonso Álamo, M. (2018). “Culpabilidad en cabeza propia y el enigma de la libertad de la voluntad”. *Cuadernos de Política Criminal*, 124, 6-7.
- Anagnostopoulos, A et al. (2015). “Viral Misinformation: The Role of Homophily and Polarization”. *WWW 2015 Companion*, 355-356.
- Attnave, F. (1959). *Applications of information theory to psychology: a summary of basic concepts, methods and results*. Oxford: Holt.
- Baldwin-Philippi, J. (2019). “The technological performance of populism”. *New Media & Society*, 21, 376-397.

- Beck, U. (1998). *La sociedad del riesgo: hacia una nueva modernidad*. Barcelona: Paidós.
- Bessi *et al.* (2014). “Social Determinants of Content Selection in the Age of (Mis)Information”. *Social Informatics SocInfo*, 8851, 259-268.
- Block, N. (1980). “What is Functionalism?” En *Readings in Philosophy of Psychology*, 1, Cambridge: Harvard University Press.
- Block, N. (1995). “Functionalism (2)”. *A Companion to the Philosophy of Mind*. Oxford: Blackwell.
- Borreguero, E. (2018). *Frente a la desinformación*. Recuperado de: [https://elpais.com/elpais/2018/12/24/opinion/1545645156\\_031421.html](https://elpais.com/elpais/2018/12/24/opinion/1545645156_031421.html)
- Bostrom, N. (2014). “Superintelligence: paths, dangers, strategies”. Oxford: Oxford U. Press.
- Brown, J *et al.* (2007). “Word of mouth communication within online communities: Conceptualizing the online social network”. *Journal of interactive marketing*, 21(3), 2-20.
- Brugnoli, E. *et al.* (2019). “Recursive patterns in online echo chambers”. *Sci Rep*, 9, 1-18.
- Chan Mora, G. (2013). “Capacidad de culpabilidad penal y libertad de decisión. Acerca del debate entre las neurociencias, la filosofía de la mente y el derecho penal”. *Revista Digital de la Maestría en Ciencias Penales*, 5.
- Choi, D. *et al.* (2020). “Rumor Propagation is Amplified by Echo Chambers in Social Media”. *Sci Rep*, 10.
- Copeland, J. (1993). *Artificial Intelligence: A Philosophical Introduction*. Oxford: Blackwell.
- Cortada de Kohan, N. (2008). “Los sesgos cognitivos en la toma de decisiones”. *International Journal of Psychological Research*, 1(1).
- Cossard, A. *et al.* (2020). “Falling into the Echo Chamber: the Italian Vaccination Debate on Twitter”. *Cornell U, International AAAI Conference on Web and Social Media*.
- Cota, W. *et al.* (2019). “Quantifying echo chamber effects in information spreading over political communication networks”. *EPJ Data Science*, 8, 1-13.
- Damasio, A. (2006). *El error de Descartes*. Barcelona: Crítica.
- De la Rosa Rodríguez, L. y Sandoval Navarro, V. (2016). “Los sesgos cognitivos y su influjo en la decisión judicial”. *Revista de Derecho Penal y Criminología*, 37, (102).
- Del Rosal Blasco, B. (2009). “La estrategia actuarial del control del riesgo en la política criminal y en el Derecho”. En Carbonell Mateu, J.C. *et al.*

- (dirs.), *Constitución, Derechos Fundamentales y Sistema Penal: (semblanzas y estudios con motivo del setenta aniversario del profesor Tomás Salvador Vives Antón)*. I. Delgado de Molina Rius, A. (2020). “Inteligencia artificial, machine learning y deep learning”. En Gurrea Martínez, A. y Remolina, N. (dirs.), *Fintech, Regtech y Legaltech: fundamentos y desafíos regulatorios*. Valencia: Tirant lo Blanch.
- Demetrio Crespo, E. (2014). “¿Libertad versus determinismo en Derecho penal?” *Revista de derecho y genoma humano: genética, biotecnología y medicina avanzada*. 1. 111-124.
- \_\_\_\_\_. “Libertad de voluntad, investigación sobre el cerebro y responsabilidad penal: aproximaciones al moderno debate sobre neurociencias y Derecho penal”. *Revista Penal México*. 6.
- Dennett, D. (1988). “Conditions of Personhood”. En Goodman, E. (ed.), *What Is a Person? Contemporary Issues in Biomedicine, Ethics, and Society*. New Jersey: Human Press.
- Díaz Arana, A. (2016). “Las mentes libres en el Derecho penal”. *InDret: Revista para el Análisis del Derecho*, 1, 1-55.
- \_\_\_\_\_. (2017). “La previsibilidad individual como fundamento de la imprudencia”. En Posada Maya, R. et al. (coords.), *Estudios Críticos de Jurisprudencia de la Corte Suprema de Justicia* 6. Bogotá: Grupo Editorial Ibáñez.
- \_\_\_\_\_. (2014). “Sobre la relevancia de los llamados ‘conocimientos especiales’ en la violación al deber objetivo de cuidado”. *Revista Libertas*. 2. Fundación Internacional de Ciencias Penales.
- Díaz Chunga, S. (2020). “Libre albedrío y constitucionalidad de la responsabilidad penal”. *Gaceta Penal & Procesal Penal*. 130, 219-239.
- Dolm, N. (2019). “Crecimiento del 713%: tecnología legal marca un récord de inversión en 2018” (traducción libre). Recuperado de: <https://www.forbes.com/sites/valentinpivovarov/2019/01/15/legaltechinvestment2018/#379851e27c2b>.
- Feijoo Sánchez, B. (2011). “Derecho Penal y Neurociencias: ¿una relación tormentosa?” *InDret: Revista para el Análisis del Derecho*. 2, 1-58.
- \_\_\_\_\_. (2012). “La culpabilidad jurídico-penal en el Estado de Derecho”. *Anuario de Derecho Penal y Ciencias Penales*. LXV, 99-125.
- \_\_\_\_\_. (2013). “Culpabilidad jurídico-penal y neurociencias”. En Demetrio Crespo, E. (dir.), *Neurociencias y derecho penal: nuevas perspectivas en el ámbito de la culpabilidad y tratamiento jurídico-penal de la peligrosidad. Segunda parte: La culpabilidad: perspectivas actuales*. Madrid: Edisofer.

- \_\_\_\_\_. (2018). *Derecho penal, neurociencias y bien jurídico*. Santiago de Chile: Olejnik.
- Fisher, M y Taub, A. (2019). “How YouTube Radicalized Brazil”. Recuperado de: <https://www.nytimes.com/2019/08/11/world/americas/youtube-brazil.html>
- Frank, R. (2002). *Sobre la estructura del concepto de culpabilidad*. Buenos Aires: B de f.
- Freeman, M y Goodenough, O. (2009). *Law, Mind and Brain*. United Kingdom: Ashgate Publishing.
- Garavito, C. (2020). “La paz y las cámaras de eco”. Recuperado de: <https://www.dejusticia.org/column/la-paz-y-las-camaras-de-eco/>.
- Garret, R. et al. (2013). “The promise and peril of real-time corrections to political misperceptions”. *Proceedings of the 2013 conference on Computer supported cooperative work*, 13, 1047-1058.
- Ginther, M. et al. (2014). “The Language of Mens Rea”. *Vanderbilt Law Review*.
- Gómez Pavajeau, C. (2018). *Neuroderecho penal y disciplinario: conducta humana, consciencia de ilicitud y reproche jurídico-social*. Bogotá: Universidad del Externado de Colombia.
- Grace, K. et al. (2018). “Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts”. *Journal of Artificial Intelligence Research*, 62, 729-754.
- Hassemer, W. (1992). “Rasgos y crisis del Derecho Penal moderno”. *Anuario de Derecho Penal y Ciencias Penales*. 45, (1).
- Hassemer, W. (1982). “¿Alternativas al principio de culpabilidad?” *Cuadernos de política criminal*, 18, 473-482.
- Hernández Giménez, M. (2019). “Inteligencia Artificial y Derecho penal”. *Actualidad Jurídica Iberoamericana*, 10 bis.
- Hilbert, M et al. (2018). “Communicating with algorithms: A Transfer Entropy Analysis of Emotions-Based Escapes from Online Echo Chambers”. *Communication Methods and Measures*, 260-275.
- Jakobs, G. (1989). “Sobre la función de la parte subjetiva del delito en Derecho penal”. Trad. Joaquín Cuello Contreras. *Anuario de Derecho Penal y Ciencias Penales*. 42(2), 633-652.
- \_\_\_\_\_. (1992). “Sobre el tratamiento de las alteraciones volitivas y cognitivas”. *Anuario de Derecho Penal y Ciencias Penales*. 45, (1), 213-234.
- \_\_\_\_\_. (1998). “La denominada actio libera in causa”. *Revista del Poder Judicial*. 50, 241-260.

- \_\_\_\_ (1998). “Conocimiento y desconocimiento de la norma”. *Problemas capitales del Derecho penal moderno: Libro homenaje a Hans Welzel*. Buenos Aires: Hammurabi.
- \_\_\_\_ (1999). “Culpabilidad y prevención”. Trad. Carlos J. Suárez. *Estudios de Derecho Penal*. Madrid: Civitas.
- \_\_\_\_ (1999). “El principio de culpabilidad”. *Estudios de Derecho Penal*. 20, Madrid: Civitas.
- \_\_\_\_ (2008). “La culpabilidad de los foráneos”. Trad. Jorge Perdomo. En Cancio Meliá, M. y Feijoo Sánchez, B. (eds.), *Teoría funcional de la pena y de la culpabilidad. Seminario con Günther Jakobs en la UAM*. Madrid: Thomson.
- \_\_\_\_ (2008). “Individuo y persona. Sobre la imputación jurídico-penal y los resultados de la moderna investigación neurológica”. En Cancio Meliá, M. y Feijoo Sánchez, B. (eds.), *Teoría funcional de la pena y de la culpabilidad: seminario con Günther Jakobs en la UAM*. Madrid: Thomson.
- \_\_\_\_ (2009). “Dolus malus”. Trad. Yamila Fakhouri. *InDret. Revista para el Análisis del Derecho*. 4, 1-23.
- Jescheck, H y Weigend, T. (1993). *Tratado de Derecho penal. Parte General*. 4a ed. Granada: Comares.
- Jones, O. et al. (2011). “Brain scans as evidence: truth, proofs, lies, and lessons”. *Mercer Law Review*, 62, 861-883.
- Jones, O. et al. (2013). “Neuroscientists in court”. *Nature Reviews Neuroscience*. 14-17.
- Jordan, G y Shorland, A. (2019). “The weekly. E. 9: ‘The Rabbit Hole’. What is. YouTube Pushing You to Watch Next?” Recuperado de: <https://www.nytimes.com/2019/08/09/the-weekly/youtube-brazil-far-right.html>.
- Kaufmann, A. (2006). *Dogmática de los delitos de omisión*. Madrid: Marcial Pons.
- Kurzweil, R. (2012). *How to Create a Mind: The Secret of Human Thought Revealed*. Nueva York: Penguin Books.
- Legg, S. y Hutter, M. (2007). “A Collection of Definitions of Intelligence”. En Goertzel, B. (ed.), *Advances in artificial general intelligence: concepts, architectures and algorithms*. IOS Press.
- López de Mántaras, R. (2019). “El futuro de la IA: hacia inteligencias artificiales realmente inteligentes”. En *¿Hacia una nueva Ilustración? Una década trascendente*. Recuperado de: <https://www.bbvaopenmind.com/wp-content/uploads/2019/02/BBVA-OpenMind-libro-2019-Hacia-una-nueva-Ilustracion-una-decada-trascendente-1.pdf>

- Luzón Peña, D. (1978). *Aspectos esenciales de la legítima defensa*. Barcelona: Bosch.
- Macadams, R y Ulen, T. (2008). “Behavioral Criminal Law and Economics”. *University of Illinois Law and Economics*, 244.
- Mackay, D. (2003). *Information Theory, Inference and Learning Algorithms*. New York: Cambridge University Press.
- Maurach, R. (1962). *Tratado de Derecho penal*. Barcelona: Ariel.
- Minsky, M. (1968). “Introduction. Artificial Intelligence and its Cybernetic Background”. *Semantic Information Processing*. Cambridge: MIT Press.
- Mir Puig, S. (2016). *Derecho penal: parte general*. 9a. ed. Barcelona: Reppertor.
- Mocanu, et al. (2015). “Collective attention in the age of (mis)information”. *Computers in Human Behaviour*.
- Moore, S. (2019). “How social media echo chambers fueled the rise of Boris Johnson”. *The Guardian*. Recuperado de: <https://www.theguardian.com/commentisfree/2019/jul/23/how-social-media-echo-chambers-fuelled-the-rise-of-boris-johnson>
- Morse, S y Roskies, A. (2013). *A Primer on Criminal Law and Neuroscience: A contribution of the Law and Neuroscience Project, support by the MacArthur Foundation*. Oxford: Oxford University Press.
- Muñoz Conde, F. (2010). *Derecho penal: parte general*. 8a ed. Valencia: Tirant lo Blanch.
- Nieva Fenoll, J. (2016). “Neurociencia y juicio jurisdiccional: pasado y presente. ¿Futuro?” *Civil Procedure Review*. 7(3).
- \_\_\_\_ (2018). *Inteligencia artificial y proceso judicial*. Madrid: Marcial Pons.
- Ortiz, R. et al. (2019). “A systematic literature review to examine the potential for social media to impact HPV vaccine uptake and awareness, knowledge, and attitudes about HPV and HPV vaccination”. *Human Vaccines & Immunotherapeutics*. 43, 1465–1475.
- Pardo, M. y Patterson, D. (2011). “Fundamentos filosóficos del Derecho y la neurociencia”. *InDret: Revista para el Análisis del Derecho*. 2, 1-51.
- Parker, J. (1993). “The Economics of Mens Rea”. *Virginia L. Review*, 79 (4), 741-811.
- Patrick, G. T. W. (1913). “The New Optimism”. *The Popular Science Monthly*, 82.
- Pérez Manzano, M. (1986). *Culpabilidad y prevención: las teorías de la prevención general positiva en la fundamentación de la imputación subjetiva y de la pena*. Madrid: Ediciones UAM.
- \_\_\_\_ (2013). “El tiempo de la conciencia y la libertad de decisión: bases para una reflexión sobre Neurociencia y responsabilidad penal”. En

Demetrio Crespo, E. (Dir.), *Neurociencias y derecho penal: nuevas perspectivas en el ámbito de la culpabilidad y tratamiento jurídico-penal de la peligrosidad. Primera parte: Libertad de acción: aspectos filosóficos, jurídicos y neurobiológicos*. Madrid: Edisofer.

Pinker, S. (2011). *Cómo funciona la mente*. Barcelona: Destino.

Quattrocchi, W. (2019). “How does misinformation spread online?” *World Economic Forum*. Recuperado de: <https://www.weforum.org/agenda/2016/01/q-a-walter-quattrocchi-digital-wildfires/>

Reyes Alvarado, Y. (2005). “El concepto de imputación objetiva”. En Bolaños G., M. (comp.), *Imputación objetiva y dogmática penal*. Bogotá: Uniandes.

Roxin, C. (1997). *Derecho Penal. Parte General*. T. I. Madrid: Civitas.

Saavedra Velasco, R. (2011). “Deconstruyendo al homo oeconomicus”. *Ius et Veritas*. 42.

Sánchez Del Campo, A. (2020). “Aspectos legales de la robótica”. En Gurrea Martínez, A. y Remolina, N. (dirs.), *Fintech, Regtech y Legaltech: Fundamentos y desafíos regulatorios*. Valencia: Tirant lo Blanch.

Sánchez Vilanova, M. (2019) *¿Neuroimputabilidad? Una mirada interdisciplinar a la responsabilidad de los trastornos de la personalidad desde los avances de las neurociencias*. Valencia: Tirant Lo Blanch.

Searle, J. (1998). *Mind, Language and Society: Philosophy in the Real World*. Nueva York: Basic Books.

Silva, J. y Varela, L. (2013). “Responsabilidades individuales en estructuras de empresa: la influencia de sesgos cognitivos y dinámicas de grupo”. En Silva Sánchez, J.M. (Dir.), *Criminalidad de empresa y compliance. Prevención y reacciones corporativas*. Barcelona: Atelier.

Silva, J. (2016). *Fundamentos del Derecho penal de la Empresa*. 2a. ed. Madrid: Edisofer.

Smith, J. (2020). “Self-consciousness. The Stanford Encyclopedia of Philosophy”. Recuperado de: <https://plato.stanford.edu/entries/self-consciousness/>

Smolensky, P. (1994). “Constituent structure and explanation in an integrated connectionist/symbolic cognitive architecture”. En Macdonald, C. y Macdonald, G. (eds.), *The Philosophy of Psychology: Debates on Psychological Explanation*. Oxford: Blackwell Publishers.

Stelter, B. (2020). “How the pro-Trump echo chamber has changed impeachment”. Recuperado de: <https://edition.cnn.com/2020/01/16/media/donald-trump-feedback-loop-reliable-sources/index.html>

- Thierer, A. y Marcus, A. (2016). “Guns, Limbs and Toys: What Future for 3D Printing?” *Minnesota Journal of Law, Science and Technology*. 17 (2).
- Turing, A. (1950). “Computing machinery and intelligence”. *Mind*, 59.
- Vives Antón, T. (2013). “Neurociencia y determinismo reduccionista: una aproximación crítica”. En Demetrio, E. (Dir.), *Neurociencias y derecho penal: nuevas perspectivas en el ámbito de la culpabilidad y tratamiento jurídico-penal de la peligrosidad. Primera parte: Libertad de acción: aspectos filosóficos, jurídicos y neurobiológicos*. Madrid: Edisofer.
- Vives Antón, T. (2002). “El principio de culpabilidad”. En Díez Ripollés, J.L. (coord.), *La ciencia del Derecho Penal ante el nuevo siglo: libro homenaje al profesor doctor don José Cerezo Mir*. Madrid: Tecnos.
- Von Belling, E. (2002). *Esquema de Derecho Penal. La Doctrina del Delito-Tipo*. Buenos Aires: El Foro.
- Welzel, H. (2006). *El nuevo sistema del Derecho Penal. Una introducción a la doctrina de la acción finalista*. Montevideo: B de f.
- Welzel, H. (1997). *Derecho penal alemán. Parte General*, 11a. ed. Santiago de Chile: Editorial Jurídica de Chile.
- Zollo, F. et al. (2015). “Emotional dynamics in the age of misinformation”. *PLoS ONE*, 10, 1-22.

