Trilogía
CIENCIA TECNOLOGÍA SOCIEDAD

Artículo de reflexión

Conducta agencial, determinismo y artefactos tecnológicos*

Agentic Behavior, Determinism, and Technological Artifacts

Daniel Augusto Duarte Arias[1]

Resumen

Según los modelos genocéntricos, el gen determina diversos aspectos de los agentes, como la inteligencia y la conducta, de modo que el entorno y la crianza tendrían poca influencia en el comportamiento de los individuos. No obstante, las teorías denominadas *4E*, en especial la tesis de la mente extendida, se oponen a este reduccionismo, al considerar que el entorno abarca un conjunto de actividades humanas que incluyen creencias, normas y moral. El objetivo de este artículo es mostrar que incluso adoptando la tesis de la mente extendida existen casos en los que la conducta humana podría estar determinada por la interacción de los agentes con artefactos tecnológicos. La tesis que sostengo es que, en algunos casos particulares, el determinismo artefactual puede darse debido a que los agentes se ven restringidos a ciertas conductas, asumiendo un rol pasivo, mientras que los artefactos desempeñan un papel mucho más activo y autónomo. Para desarrollar este propósito, primero examinaré qué es la agencia y cómo el determinismo genético ofrece un marco explicativo sobre la conducta humana. Luego, analizaré el externalismo activo, la tesis de la mente extendida y el concepto de transparencia como un marco explicativo opuesto al determinismo genético. Finalmente, argumentaré cómo, en ciertos escenarios de determinismo artefactual, pueden surgir casos ambiguos de conducta humana debido a los *affordances* de los artefactos tecnológicos y a una serie de criterios adicionales.

Palabras clave: artefacto tecnológico, conducta agencial, determinismo, externalismo activo, mente extendida.

Cómo referenciar

Duarte Arias, D. A. (2025). Conducta agencial, determinismo y artefactos tecnológicos. *Trilogía Ciencia Tecnología Sociedad*, *17*(35), e3300. https://doi.org/10.22430/21457778.3300

* Producto desarrollado como estudiante en formación de doctorado con el proyecto de investigación *El transhumanismo y la tesis de la mente extendida: ¿puede la tecnología transformar radicalmente la cognición humana?* (código 2023-59530), liderado por el profesor Ángel Rivera-Novoa, financiado por el Comité para el Desarrollo de la Investigación (CODI) de la Universidad de Antioquia.
[1] Universidad de San Buenaventura, Cali, Colombia, daduartea@usbcali.edu.co

Recibido: 28 de octubre de 2024
Aceptado: 06 de marzo de 2025

CC BY NC SA
OPEN ACCESS

Abstract
Gene-centered models hold that genes dictate various aspects of agents, such as intelligence and behavior, thereby minimizing the role of the environment and upbringing in shaping individual conduct. However, 4E theories, especially the extended mind thesis, challenge this reductionist perspective by emphasizing that the environment encompasses a set of human activities, including beliefs, norms, and morals. Against this backdrop, this paper aims to demonstrate that, even through the lens of the extended mind thesis, there are instances in which human behavior remains determined by interactions with technological artifacts. Specifically, it is argued that artifactual determinism may occur in situations where agents are restricted to particular behaviors, assuming a passive role, while technological artifacts take on a much more active and autonomous role. To develop this argument, the discussion begins by examining the concept of agency and how genetic determinism offers an explanatory framework for human behavior. Following this, it analyzes active externalism, the extended mind thesis, and the notion of transparency as an alternative paradigm to genetic determinism. Finally, the article contends that, in certain scenarios of artifactual determinism, ambiguous forms of human behavior may arise due to the affordances of technological artifacts and other factors.

Keywords: technological artifact, agentic behavior, determinism, active externalism, extended mind.

INTRODUCCIÓN

La ley genética del comportamiento es un modelo explicativo de la conducta basado en la idea de que nuestra naturaleza y acciones están determinadas por factores genéticos. Al recurrir al ácido desoxirribonucleico (ADN) como único nivel explicativo se incurre en un determinismo genético moral. Esta afirmación también se conoce como esencialismo genético. Lippman (1993) lo define como:

> ... el proceso continuo por el que se da prioridad a las diferencias entre individuos basadas en sus códigos de ADN, con la mayoría de los trastornos, comportamientos y variaciones fisiológicas (incluidas cosas como la esquizofrenia y la hipertensión arterial, así como la afinación perfecta y la capacidad de los niños para sentarse quietos a ver la televisión) estructurados como, al menos en parte, hereditarios. (p. 178)[1]

De ser cierto este reduccionismo, los rasgos comportamentales estarían exclusivamente involucrados con el ADN y constituirían a los seres humanos. Según Turkheimer (2000) esto obedece a tres leyes que determinan el comportamiento humano porque: 1) el comportamiento es heredable; 2) la crianza ejerce menor influencia que los genes y 3) los efectos genéticos y de crianza no explican una parte sustancial de los cambios comportamentales.

Las ideas de la genetización se refuerzan con tesis y argumentos sostenidos por Dawkins (2006), Nelkin y Lindee (1995), porque defienden la jerarquía del gen sobre el comportamiento humano.

[1] En adelante las traducciones de la lengua original al español son propias del autor de este artículo.

Sin embargo, para Sharon (2013) este tipo de determinismo iría en contra de explicaciones conductuales mucho más robustas, ya que ignoran la incidencia del ambiente y sus repercusiones en el sujeto.

Así las cosas, el problema radica en que, para los esencialistas genéticos, es irrelevante cuánto incida la crianza, como factor externo, en la conducta humana, pues solo sería posible un comportamiento distinto mediante la transformación o modificación de los genes. Los argumentos teóricos de las ciencias cognitivas *4E*, que conciben la cognición como *embodied* (influenciada por las características físicas), *embedded* (anclada especialmente en el entorno), *extended* (ampliada mediante herramientas), *enactive* (emergente de la interacción agente-entorno) (Shapiro y Spaulding, 2024), hablarían en contra de este determinismo, pues la tesis de la cognición distribuida propone que hay un acople entre los agentes y el entorno, dando lugar a que aspectos como la crianza incidan en la conducta humana. Especialmente me centraré en la tesis de la mente extendida que se opone al intracranealismo porque defiende que la mente (procesos cognitivos, deseos, recuerdos, pensamientos) se extiende más allá de los límites del agente biológico (Clark y Chalmers, 1998).

El objetivo de este artículo es mostrar que, incluso adoptando la tesis de la mente extendida, la conducta humana podría estar determinada por la transparencia y los *affordances*[2] de los artefactos con los que el agente interactúa. Lo anterior implicaría que en algunos contextos se incurra en un determinismo artefactual, es decir, en que necesariamente los artefactos tecnológicos determinan los modos de acción de agentes biológicos, en especial con artefactos computacionales autónomos (Turilli, 2007; Turilli y Floridi, 2009), debido a la necesidad técnica que emerge de la propia naturaleza del sistema tecnológico.

La tesis que sostengo es que en algunos casos particulares los agentes se ven restringidos a ciertas conductas, lo que implicaría que en un sistema de mente extendida los artefactos desempeñan un papel mucho más activo y los agentes asumirían un rol pasivo. Por ejemplo, aunque un martillo permita algunos usos, por su modelación técnica está limitado a ser una extensión física de capacidades mecánicas específicas. En cambio, algunos artefactos computacionales autónomos están diseñados específicamente para emular y realizar funciones humanas. Pueden procesar información, aprender de la experiencia, tomar decisiones basadas en criterios complejos y adaptarse a nuevas situaciones. En consecuencia, en algunos casos la responsabilidad sobre la conducta humana puede ser ambigua por la pérdida de autonomía del agente biológico.

Para llevar a cabo este propósito, primero examinaré qué es la agencia y cómo el determinismo genético ofrece un marco explicativo sobre la conducta humana. Luego, analizaré el externalismo activo, la tesis de la mente extendida y el concepto de transparencia como un marco explicativo opuesto al determinismo genético. Finalmente, argumentaré cómo, a partir del determinismo artefactual, pueden darse casos ambiguos de conducta humana debido a los *affordances* de los

[2] El término *affordances* generalmente hace referencia a las posibilidades de acción que los artefactos y el entorno le proporcionan a un agente en una relación bidireccional. Más adelante se hará una ampliación de este concepto.

artefactos tecnológicos, que cumplen con criterios como la delegación, pérdida del crédito moral y la transparencia. El artículo concluirá con algunas reflexiones finales sobre los aspectos generales discutidos.

AGENCIA, CONDUCTA HUMANA Y DETERMINISMO GENÉTICO

La agencia es la capacidad que tiene un ser para actuar. A partir de las ideas de Anscombe (1991) y Davidson (1963, 1971), se estructura la concepción estándar de la acción que afirma que esta debe explicarse con base en la intencionalidad y que toda acción intencional requiere de operaciones racionales. En este sentido, hay una primacía de la acción intencional sobre la mera acción. Un ejemplo de agencia es que a partir de una descripción como «escribir en la computadora» un sujeto desarrolle esta acción que desea. Un comportamiento agencial implica no solo la realización de una acción sino además desear o tener intención de llevar a cabo una acción. No obstante, teorías clásicas del comportamiento afirman que nuestras conductas están estrechamente relacionadas con los genes, por lo cual, parte de lo que deseamos se encuentra determinado por la genética que nos constituye.

Las tesis esencialistas de la genética explican la biología mediante tres conceptos básicos: composición, estructura y dinámica. Estos conceptos mostraban cómo funcionaba el universo natural porque daban cuenta de las moléculas que contenía el organismo, cómo se distribuían las moléculas en el espacio y cómo interactuaban estas moléculas a nivel químico. Sin embargo, con la aparición de las ideas de Crick (1958), emerge un cuarto concepto (el de información) que contribuye a una explicación estrecha de la biología en general.

El razonamiento de Crick sostiene que el ADN constituye a los organismos fundamentalmente porque es el que contiene la información de diversas proteínas que predisponen la expresión genética. Así las cosas, aspectos como la conducta o la moral se ven reducidas a los genes y es mediante estos que se explican las acciones humanas. Una teoría explicativa compatible con lo anterior se denomina *síntesis moderna*. Wilson (1975) asume el marco explicativo de Darwin (1859), en la teoría de la evolución porque reduce los aspectos humanos y sociales a la *síntesis moderna* en la que «se sopesa la importancia adaptativa de cada fenómeno y se relaciona con los principios básicos de la genética de poblaciones» (p. 4). Las distintas cuestiones fisiológicas y morales obedecen a transformaciones biológicas y evolutivas que se perpetúan en el ADN. En otras palabras, las conductas agenciales estarían intrínsecamente vinculadas a los organismos, debido a su constitución genética.

Este tipo de creencias reduccionistas también vienen motivadas por otras ideas de Darwin (1872), quien argumentaba que el comportamiento no era adquirido por la crianza, sino que obedecía a factores hereditarios. Darwin refuerza esta idea con una anécdota de la esposa de un hombre prestigioso que constantemente notaba que su marido, cuando dormía en posición con la cabeza y el torso mirando hacia el techo, levantaba su brazo derecho lentamente hasta que estuviera frente a su cara para luego dejarlo caer de forma abrupta sobre su nariz. La esposa del hijo de esta pareja, quien jamás había escuchado de esta conducta del hombre prestigioso, notó que su esposo tenía

esta misma peculiaridad a la hora de dormir. También que su hija había heredado este mismo comportamiento (Darwin, 1872, pp. 33-34).

Lykken (2000) sugiere que el comportamiento, al estar determinado por factores genéticos, podría ofrecer predicciones sobre la actuación de las personas en contextos determinados. Así pues, si la información que la estructura del ADN proporciona a los científicos muestra rasgos o patologías asociadas a comportamientos criminales, entonces podría intervenirse de forma temprana para corregir estas conductas y evitar actitudes o acciones no deseadas.

De ser ciertas las ideas de Wilson y Lykken el determinismo implicaría afirmaciones más problemáticas como la idea del genocentrismo, para el cual los rasgos comportamentales están limitados dentro de canales genéticos estrechos. Rosoff y Rosenberg (2006) afirman que el genocentrismo es:

> ... una afirmación sobre los orígenes y la invariabilidad, más allá de la intervención humana, de las capacidades, disposiciones y comportamientos de las personas que tienen consecuencias jurídicas, económicas, políticas, interpersonales y sanitarias para estas personas y para otras con las que interactúan. (p. 127)

En consecuencia, el comportamiento de los agentes estaría determinado por los genes que, a su vez, limitarían en gran medida la influencia del entorno.

La genetización se convierte también en un marco explicativo para niveles tecnológicos como la ingeniería genética, pues con esta herramienta se pueden mejorar, inhibir o eliminar conductas no deseables. Persson y Savulescu (2008) proponen que mediante técnicas asociadas a la ingeniería genética se podrían mejorar aspectos comportamentales del ser humano debido a la alta influencia de los genes sobre la conducta humana. Si bien la crianza o el entorno pueden modelar ciertas conductas, los genes determinan los principales aspectos del comportamiento humano, por lo cual es en este nivel genético donde pueden darse las modificaciones y mejoras conductuales (p. 168).

Lewontin et al. (1984) argumentan que este biologismo sostiene que el comportamiento puede ser modificado gracias a que la conducta se encuentra localizada en el cerebro y solo bastaría con una intervención en los genes no deseados para transformar las propiedades conductuales anómalas del ser humano (p. 57). Esta perspectiva daría cuenta de que los comportamientos del ser humano también se asocian a rasgos de la inteligencia, por lo cual intervenir el cerebro haría que algunos procesos cognitivos (como los asociados al razonamiento) contribuyeran a mejorar el comportamiento de los individuos.

Las visiones unidireccionales de la biología intentan sostener la idea de que solo basta con la modificación molecular de los genes para transformar la sociedad. Dawkins (2006) asume su compromiso con el dogma central de la biología molecular porque en sus ideas afirma que el gen es algo previo al individuo, y la sociedad se constituye por los individuos. Por tal razón, la conducta

misma de una sociedad solo podría verse afectada si a nivel molecular ocurre un cambio que afecte el comportamiento del individuo.

Un ejemplo ilustrativo sobre estas ideas reduccionistas de la conducta al nivel genético biológico podría ser un individuo que constantemente recae en acciones criminales. Para los defensores del biologicismo, esta conducta se explicaría con base en la información genética del individuo. También podría afirmarse que estas conductas serían heredadas principalmente porque están contenidas en la información del ADN del individuo. En consecuencia, la predicción sobre la descendencia de este sujeto es que sus herederos genéticos serán propensos a conductas criminales. Una forma de evitar estas conductas sería implementar la eugenesia o un proyecto de cría para evitar que esta información se propague a otros organismos.

En general, el determinismo genético estaría en contra de explicaciones sobre la conducta humana que involucren el entorno, puesto que es en la información genética donde se encuentra la manera de contrarrestar algunas conductas no deseables o impulsar aquellas que se considerarían deseables para los seres humanos. Por otra parte, la naturaleza humana podría transformarse de tal manera que los comportamientos deseables puedan heredarse mejorando los comportamientos de una sociedad de forma unidireccional.

Así pues, si el determinismo biológico sostiene que el gen constituye varios aspectos de los agentes como la inteligencia y la conducta, entonces el entorno y la crianza tendrían poca influencia en el comportamiento de los individuos. Este reduccionismo estaría en contra de tesis cognitivistas como la mente extendida que, a diferencia de la visión determinista de la genética, afirma que el entorno puede incidir de forma activa en rasgos como la inteligencia y por tanto habría algunos procesos mentales que podrían ocurrir fuera del cerebro.

EXTERNALISMO ACTIVO, MENTE EXTENDIDA Y TRANSPARENCIA

Si bien el determinismo genético puede explicar algunos comportamientos humanos, dichos comportamientos no se reducen exclusivamente al nivel de los genes. El entorno comprende un conjunto de actividades humanas que incluye creencias, normas y moral. Estas actividades que suponen contenidos mentales (como los deseos) dan cuenta de que el comportamiento humano involucra un proceso de inculturación donde tiene lugar un sistema en el que participan agentes, entorno y artefactos (Monterroza-Rios y Gutiérrez-Aguilar, 2022).

Las teorías denominadas *4E* se oponen a las visiones intracranealistas o reduccionistas de los procesos cognitivos y a las visiones genocéntricas. Aunque no son enfoques unificados, comúnmente se conoce como *4E* a la tesis que sostiene que «la cognición no se limitaba a procesos cerebrales, sino que estaba incorporada, integrada, extendida y enactiva» (Newen et al., 2018, p. 4). En este sentido, son explicaciones cognitivistas opuestas con coherencia a los puntos de vista exclusivamente internalistas porque el entorno juega un rol activo en la realización de, por ejemplo, procesos cognitivos.

Lo anterior es compatible con los argumentos de Varela et al. (1993), Di Paolo et al. (2017) o Fuchs (2017), quienes exponen que hay un acople entre los agentes y el entorno, dando lugar a que aspectos como la crianza incidan en la conducta humana.[3] En este caso, los procesos conductuales no solo serían explicables desde el nivel genético, sino que también ocurren y se sitúan en la interacción con el entorno. De hecho, las relaciones sociales darían lugar a que se entrelacen distintos procesos que moldeen la conducta o el comportamiento humano. Monterroza-Rios (2023) afirma que esto es posible porque: 1) los organismos no solo se adaptan al entorno sino que también lo modifican (construcción de nichos) (Laland et al., 2000) y; 2) ocurre el efecto *Ratcher* que es una mezcla entre la herencia biológica y cultural que perdura en una memoria colectiva como los artefactos o las instituciones (Tomasello, 2003). De cumplirse estos dos criterios, podría decirse que en un entorno de prácticas en el que actúan diversos agentes y artefactos ocurre un enactivismo cultural (Monterroza-Rios, 2023, p. 228).

La tesis de la mente extendida explica también cómo ocurren o no procesos mentales fuera del cráneo y, por tanto, parte de nuestros comportamientos no están exclusivamente asociados al cerebro. Esta tesis es una perspectiva cognitivista que sostiene que los artefactos tecnológicos podrían considerarse parte constitutiva de nuestra mente. Kiverstein (2018) lo expone de la siguiente manera:

> Las acciones corporales y los recursos del entorno sobre los que actúan los agentes pueden, en determinadas condiciones, contar como partes constituyentes de un proceso cognitivo. Pensemos, por ejemplo, en lo integrados que están los teléfonos móviles en esos momentos de nuestras vidas en los que nos quedamos con nuestros propios pensamientos [...] Los teléfonos inteligentes y otras tecnologías móviles están tan profundamente vinculados en nuestra vida cotidiana que, según [la tesis de la mente extendida], ahora se podría pensar en ellos como partes de nuestras mentes. (p. 19)

Estas ideas están motivadas por un ensayo clásico de Clark y Chalmers (1998), en el que los autores muestran cómo algunos contextos pueden dar cuenta de la extensión de la mente mediante el uso de artefactos que permitirían la realización de procesos cognitivos fuera del cerebro. Clark y Chalmers (1998) basan esta tesis en el principio de paridad en el que afirman que:

> Si al enfrentarnos a alguna tarea, una parte del mundo funciona como un proceso que, si se desarrollara en la cabeza, no dudaríamos en aceptar como parte del proceso cognitivo, entonces esa parte del mundo es (por ese momento) parte del proceso cognitivo. (p. 8)

Un ejemplo de cómo pueden ocurrir estos procesos extendidos es el uso de artefactos como el lápiz y papel, una calculadora o un dispositivo electrónico que permita realizar operaciones matemáticas. Un individuo podría recurrir solo a su cerebro para llevar a cabo operaciones matemáticas que dispongan de procesos como el razonamiento, la memoria o la atención. La tesis de la mente extendida sostendría que el individuo podría realizar también estos procesos, incluso aprovechar algunas ventajas de los artefactos como el lápiz y papel, para que estas tareas sean realizadas con

[3] Para Monterroza-Rios, el cuerpo, los sistemas metabólicos y fisiológicos se encuentran situados en interacción con el mundo, lo que incluye las relaciones sociales y los hábitos (Monterroza-Rios, 2023, p. 223).

apoyo del entorno. Podrían darse casos en los que artefactos como la calculadora o dispositivos electrónicos apoyarían estas tareas cognitivas si cumplen con el principio de paridad propuesto por Clark y Chalmers.

Sin embargo, Clark y Chalmers (1998) proponen un razonamiento más radical. Los autores sostienen que incluso es posible extender creencias y, por tanto, esto daría cuenta de la extensión de la mente:

> Proponemos ir un paso más allá. Aunque algunos estados mentales, como las experiencias, pueden determinarse internamente, hay otros casos en los que los factores externos contribuyen de forma significativa. En concreto, argumentaremos que las creencias pueden estar constituidas en parte por características del entorno, cuando esas características desempeñan el papel adecuado en el impulso de los procesos cognitivos. Si es así, la mente se extiende al mundo. (p. 12)

Los procesos cognitivos que son impulsados por algunas creencias podrían también contribuir a la realización de acciones en la vida cotidiana. Uno de los vehículos que permitiría la extensión de la mente, según Clark y Chalmers, es el lenguaje. Con el fin de ilustrar esta extensión recurren al ejemplo de un individuo (Otto) que padece de Alzheimer. Otto confía en el entorno para estructurar su vida y, para este apoyo, utiliza un ordenador portátil donde almacena información que posteriormente podrá ser consultada.

De este modo, el portátil cumple en ocasiones la función de memorizar, por ejemplo, direcciones de lugares que Otto requiere para su vida cotidiana pues «Otto utiliza constantemente su *notebook*. Es fundamental para sus acciones en todo tipo de contextos, del mismo modo que una memoria ordinaria es fundamental en una vida ordinaria» (Clark y Chalmers, 1998, p. 13). A pesar de que la tesis de la mente extendida inicialmente puede asociarse al cognitivismo, un aspecto relevante del ejemplo de Otto es que la interacción con el entorno le es útil para aspectos de su vida cotidiana como las acciones que puede realizar por el procesamiento de información.

Esta primera formulación de la tesis de la mente extendida es conocida como la «primera ola». No obstante, han surgido otras versiones de esta tesis con el fin de explorar las diferentes interacciones entre artefactos e individuos dentro de un sistema. Por ejemplo, la «segunda ola» desarrollada por Sutton (2010) o Menary (2007, 2010) sostiene que en lugar del principio de paridad se debería explorar la complementariedad o la manipulación de los artefactos en los procesos cognitivos. La «tercera ola» radicaliza aún más la tesis de la mente extendida, enfatizando en que los límites de la mente deben ser renegociados constantemente debido a la interacción de los agentes con el entorno (Kirchhoff y Kiverstein, 2019).

Autores como Andrada et al. (2023) desarrollan el concepto de *transparencia* para explicar cómo ocurren casos de mente extendida que guían la cotidianidad de los agentes. En las conductas agenciales no solo se encuentran involucradas las creencias, sino que los artefactos o el entorno cobran un papel activo. Para explicar esta relación, la transparencia se clasifica en dos perspectivas. La primera perspectiva se define como transparencia reflexiva:

> ... es decir, ver distintos aspectos del equipamiento y los mecanismos del sistema de IA con el que interactúan los seres humanos) se refiere a aquellos aspectos de la agencia humana por los que somos capaces de observar aspectos del funcionamiento de un artefacto, aplicación o sistema de software determinado. (p. 1327)

La segunda perspectiva es la transparencia en el uso que «resalta la capacidad de un agente para mediar algún aspecto de su acción hacia un objetivo determinado de manera controlada, fluida y hábil» (Andrada et al., 2023, p. 1327). Esta perspectiva tiene antecedentes en las secciones § 15-16 de *Ser y tiempo* de Heidegger pues los entes a la mano (*Zuhandenheit*) se hacen «transparentes» mientras funcionan adecuadamente en el flujo de nuestras actividades cotidianas (Heidegger, 2016, pp. 89-98). Así las cosas, el primer tipo de transparencia se refiere a la capacidad de ver y comprender los componentes internos y efectos de un artefacto, mientras que el segundo es la capacidad de un agente para actuar a través de una herramienta o artefacto de manera fluida, sin necesidad de reflexión consciente sobre su uso.

La tensión entre ambas perspectivas se da porque la transparencia reflexiva requiere constantemente atención y consciencia por parte del agente biológico, mientras la transparencia en el uso requiere que la consciencia y la atención se alejen de la manipulación del artefacto y más bien se sitúen en la tarea que se desea realizar. Podrían ocurrir contextos en que la transparencia de uso obvie la manera en que algunos artefactos tecnológicos funcionan porque lo que requiere su atención es la tarea que se desempeña con el uso de algunas herramientas. Así pues, mientras se realizan tareas o se llevan a cabo acciones en interacción con entornos, los artefactos tecnológicos autónomos podrían llegar a sustituir a los agentes biológicos en un sistema de mente extendida. De ser cierto este razonamiento, algunos casos de uso de herramientas podrían conducir a ambigüedades sobre cuándo hay o no autonomía, responsabilidad o determinismo en las conductas de un agente.

CONDUCTA AGENCIAL Y DETERMINISMO ARTEFACTUAL

Hasta el momento, en el desarrollo argumentativo de este artículo se ha dicho que la tesis de la mente extendida supone que el uso de artefactos involucra la interacción de un agente con el entorno. Sin embargo: ¿qué son los artefactos? Pues bien, estas herramientas son:

> ... elementos materiales a los que les son moldeados las formas y la composición a través de los diseños y planes de acción (explícitos o no) de agentes técnicos. Estas piezas materiales tienen el propósito de cumplir determinadas funciones (prácticas o simbólicas) relevantes a ciertos grupos humanos, que reconocen con su uso. (Monterroza-Ríos, 2018, p. 211)

De este modo, un artefacto podría propiciar conductas en lo individual, pero también en lo social, ya que estos elementos materiales inciden y se desarrollan en redes institucionales y simbólicas que involucran diversos usos en múltiples entornos. Alrededor de esta concepción artefactual se establece que los agentes adoptan acciones según ciertas capacidades que ofrecen los elementos y el entorno. El concepto de *affordance* explicaría esta relación. Gibson (2015) afirma que:

> Las *affordances* del entorno son lo que este ofrece al animal, lo que le proporciona o suministra, ya sea para bien o para mal. El verbo *to afford* se encuentra en el diccionario, pero el sustantivo *affordance* no. Lo he inventado. Con esto quiero decir algo que se refiere tanto al entorno como al animal de una manera que ningún término existente lo hace. Implica la complementariedad entre el animal y el entorno. (p. 119)

Es así como un artefacto complementa las acciones que involucran la interacción de un agente con el entorno, de modo que pueden ocurrir diversas acciones porque los artefactos soportan y ofrecen estas posibilidades a los usuarios. Gibson da varios ejemplos de cómo los *affordances* permiten, a través de la manipulación de artefactos como cuchillos, garrotes o martillos, múltiples usos por parte de los agentes (Gibson, 2015, p. 125). Lo que estos ejemplos tienen en común es que son compatibles con el concepto de transparencia, puesto que para Gibson estas interacciones inciden en el comportamiento de los agentes al ser constitutivos en un sistema. Fasoli (2018) señala que la complejidad de los artefactos y sus diseños conduce a una visión parcialmente determinista porque es en los *affordances* donde subyace la posibilidad de la función sustitutiva de estos. Algunos artefactos, de forma intrínseca, podrían empujarnos a realizar procesos determinados (Fasoli, 2018, p. 15).

Una postura similar es la de Verbeek (2005), basado en la teoría de *actor red* de Latour (2005). Verbeek establece cuándo, en ciertas circunstancias y con ciertos criterios se pueden dar relaciones mediadas entre agentes y artefactos. El primer criterio es la traducción o transformación, por la que en contextos específicos los artefactos y agentes se ven involucrados en posibilidades de acciones. El segundo, denominado *composición*, es la asociación entre agentes y artefactos como actantes. El tercero es el *black-boxing* reversible, en el que algunas relaciones entre agentes y artefactos permanecen ocultas o transparentes. Por último, la delegación hace referencia a que se puede desplazar o trasladar la tarea de un agente a un artefacto (Verbeek, 2005, pp. 155-160).

Lo que tienen en común los razonamientos de Gibson y Verbeek es su compatibilidad con la tesis de la mente extendida. La conducta humana podría estar mediada por artefactos tecnológicos no solo debido a las características que tiene la mente, sino además porque un sistema entre agente y artefactos está mediado de tal modo que pueden llevarse a cabo tareas que incidan en la conducta humana. Las ideas antes expuestas sirven como marco explicativo del determinismo artefactual.

El determinismo artefactual que quiero argumentar es un caso particular en el que se cumplen criterios suficientes tales que:

a) Un agente delegue gran parte de sus tareas y acciones a un artefacto tecnológico autónomo.
b) Suceda un detrimento o pérdida del crédito epistémico y moral.
c) Ocurra la transparencia en el uso y haya escasa o nula reflexión, de tal forma que el agente biológico pierda su capacidad crítica.

Un contraargumento es el elaborado por Parselis (2018), según el cual los artefactos no determinan unilateralmente el comportamiento humano. Lo que ocurre es que tanto los artefactos como los

agentes se dan forma continuamente en el marco de las tecnologías, incluso de aquellas que se denominan entrañables por su modo de vinculación al agente. En este sentido, Parselis (2018) afirma que «si los artefactos son portadores de la apertura de un espacio de posibilidades, debemos poder juzgar en qué consisten las posibilidades que podemos crear» (p. 53). Sin embargo, Parselis no niega que los artefactos pueden incorporar y transmitir algunos valores que inciden en el comportamiento humano, lo que indicaría que tanto las normas como las instituciones pueden modelarse también por la interacción de los agentes con diversos elementos materiales.

Ahora bien, el determinismo artefactual, a diferencia del determinismo genético, es el que, en algunos casos, cuando hay sistemas complementarios de extensión exitosa, da lugar a la manipulación de ciertos artefactos autónomos que puede determinar la acción. En este sistema, los agentes pasivos siguen siendo parte del entramado entre el entorno activo y los artefactos que también cumplen un rol activo. Los agentes en estos contextos deterministas pierden ciertas características que conllevan la consideración de su falta de autonomía, crédito epistémico, crédito moral y capacidad crítica.

Nyholm (2024) advierte que parte de la crítica y las discusiones alrededor de la inteligencia humana radica en la pregunta de la incidencia de los artefactos tecnológicos sobre la acción humana. Particularmente, para Nyholm, el uso de inteligencias artificiales (IA) abre el debate sobre si se corren riesgos por una forma de «degradación humana». Si bien esta degradación podría ser examinada en el nivel cognitivo, mi interés en los razonamientos expuestos por Nyholm radica en que también puede ocurrir una «crisis de pasividad moral» en los agentes que involucran sus actividades cotidianas con artefactos tecnológicos (p. 78).

Nyholm (2024) ofrece una definición de inteligencia que es útil a la discusión porque involucra capacidades y procesos cognitivos con el mundo y el comportamiento. De este modo, la inteligencia debe ser entendida como:

> ... un conjunto de capacidades cognitivas que nos ayuda a ampliar nuestro conocimiento y comprensión del mundo, y que también nos ayuda a alcanzar mejor nuestros objetivos y a vivir de acuerdo con nuestros valores. Además, asumiré que la inteligencia es (i) un potencial básico que tienen los seres humanos —y que también pueden tener los miembros de otras especies, aunque de distinto tipo— y/o (ii) un potencial más o menos plenamente realizado, por lo que distintas personas pueden ser más o menos inteligentes en su pensamiento o comportamiento. (pp. 78-79)

Cabe destacar que Nyholm vincula los rasgos de un agente inteligente con una capacidad para alcanzar objetivos y vivir según ciertos valores, por lo que las capacidades de la inteligencia inciden en el pensamiento y el comportamiento. Algunas IA sofisticadas modeladas con la técnica de *machine learning* o aquellas redes neuronales programadas como modelos lingüísticos podrían propiciar, según Nyholm, algunos contextos donde los usuarios actúen como si hubieran mejorado su comportamiento. La razón de esta afirmación es que, si bien algunos usuarios pueden realizar acciones con estas tecnologías, la producción o los comportamientos que se desarrollan por la manipulación de estas máquinas no se asocia a la inteligencia humana sino a conductas y

comportamientos artificialmente inteligentes. Respecto al criterio de autonomía, podría afirmarse que en cuanto mayor autonomía tenga un artefacto tecnológico, el agente humano puede adquirir una conducta pasiva o perder autonomía frente a los comportamientos que se realicen por la complementariedad de estos artefactos en un sistema extendido. Dicho de otra manera, cuanto más delegue un agente sus comportamientos a una máquina inteligente, menor autonomía tendrá. De este modo, si ocurre una delegación total, habría una pérdida o detrimento de autonomía total.

Nyholm (2024) afirma que cuando ocurren estos comportamientos artificialmente inteligentes el agente en ciertos contextos no podría ser digno de elogio (p. 85). Para Rivera-Novoa (2024), estas conductas reflejan un tipo de ignorancia particular que se asocia a no saber *cómo es actuar* o realizar un *comportamiento como*. Para este autor «cuando hay extensión cognitiva, pero delegación total del crédito epistémico, entonces podemos ser ignorantes de cierta fenomenología cognitiva: no sabemos qué es hacer una tarea cognitiva por nuestros propios medios» (p. 14). De este modo, podríamos ser ignorantes de cómo es ser un agente que razona, comprende, piensa, entiende, etc. Sin embargo, la tesis de Rivera-Novoa puede llevarse más allá si se asocia con algunos aspectos que Nyholm relaciona con la inteligencia.

Los casos de delegación total que conducen a la ignorancia del crédito epistémico podrían ser asociadas también a la ignorancia de una especie particular de fenomenología moral. Hume (1990), en *An enquiry concerning the principles of morals*, establece en la primera sección que la norma que dicta la satisfacción por el mérito personal es el sentimiento que producen nuestras conductas porque permiten considerar:

> ... cada atributo de la mente que convierte a una persona en objeto de estima y afecto, o de odio y desprecio; cada hábito, sentimiento o facultad que, si se atribuye a alguien, implica alabanza o culpa, y puede entrar en cualquier panegírico o sátira sobre su carácter y modales. (p. 16)

Un agente que desplaza parte de su crédito moral para que un artefacto tecnológico realice ciertos comportamientos en contextos determinados no puede ser digno de elogios o méritos por la conducta. El caso es que, si esto es así, puede haber ocasiones en las que en extensiones genuinas de la mente donde se deleguen totalmente nuestras conductas, se produzca ignorancia en experiencias morales. Los agentes no sabrían *cómo es actuar* o realizar un *comportamiento como* un agente que tiene sentimientos o emociones por sus conductas, hábitos, carácter o modales.

Así pues, una consecuencia del detrimento o la pérdida del crédito epistémico y moral es que si se ignoran ciertas experiencias fenomenológicas cognitivas y morales, un agente perderá la experiencia de ser como alguien que discierne según sus sentimientos o como alguien que aprueba o desaprueba ciertas conductas guiado por sus emociones. Al no saber comportarse como un agente moral o no saber *cómo es actuar* como un agente moral, entonces habrá comportamientos en un sistema extendido que puedan determinar su comportamiento. Con lo anterior, el criterio de transparencia daría cuenta de cómo un agente podría desconocer parte de cómo se elabora un artefacto, cómo se programa o cómo realiza tareas. Mientras ocurre una transparencia en el uso, es decir, en la complementariedad entre un agente y un artefacto para desarrollar una tarea, es posible

que el agente pierda su capacidad crítica por la falta de conciencia sobre cómo desarrolla el artefacto esta actividad.

Algunos casos de extensión de la mente pueden llevar al determinismo artefactual porque los agentes se ven restringidos a conductas particulares. Los artefactos tecnológicos desempeñarían un rol mucho más activo que algunos agentes que han delegado totalmente algunas conductas y por ende habrán perdido ciertas experiencias fenoménicas morales como *saber actuar como*. Sin embargo, una consecuencia de este razonamiento es que en algunos casos la responsabilidad sobre la conducta humana pueda ser ambigua por la pérdida de autonomía del agente o la delegación total de la conducta en artefactos tecnológicos.

En la serie *Lost* (Abrams y Lindelof, 2004) puede analizarse un caso ilustrativo sobre este tipo de determinismo artefactual. Expondré dos ejemplos para examinar los criterios propuestos en este artículo. En el primer ejemplo, en la serie, la estación Cisne se encuentra en medio de una isla habitada durante los años setenta por diferentes científicos. Muchos años después, la isla estaba parcialmente inhabitada pues solo se encontraba un científico llamado Kelvin dentro de la estación. Desmond es quien llega a explorar la isla, luego encuentra la estación y conoce a Kelvin, quien le cuenta que su rutina diaria es ingresar una serie de números en una computadora cada 108 minutos para evitar una catástrofe electromagnética global, pues el laboratorio cumple la función de contener esta energía anómala. Kelvin muere debido a un accidente y Desmond asume la función de digitar el código siguiendo la misma rutina. Poco después otros personajes aparecen en la isla, encontrando también la estación y asumiendo la tarea de digitar el código cada 108 minutos. La estación tiene una especie de válvula que, de fallar el ingreso de los números, permite implosionar el laboratorio para evitar el impacto electromagnético global.

El segundo ejemplo ocurre en el mismo contexto planteado para el anterior, pero a la llegada de los demás personajes, uno de ellos logra que la computadora en la que se ingresan los números pueda programarse con una IA modelada con la técnica de *machine learning*. Con el pasar de los días la máquina aprende a controlar varios elementos del laboratorio como abrir las puertas cuando ingresa o sale alguien, encender las luces, enviar avisos sobre el contador de tiempo e incluso digitar los números antes de que acabe este contador. Al principio, los personajes vigilan las actividades que realiza la computadora con el fin de no dar lugar a errores y de vez en cuando digitar ellos los números. Cada vez más van delegando esta tarea incluso hasta el punto de no supervisar la labor porque la computadora procesa bien la información y aprendió a hacer la tarea justo como ellos lo harían. También hicieron un digitador del código manual para usarlo cuando se necesitara. La máquina inteligente también aprendió que la válvula puede ser utilizada en caso de emergencia, pero añadió en su sistema la función de activar la válvula en caso de alguna anomalía, con un breve temporizador, si perdiera el control de la estación y no hubiera peligro de que por la implosión ocurrieran decesos. Un día cualquiera, los habitantes de la isla se encuentran lejos del laboratorio realizando labores de recolección, pero debido a una tormenta eléctrica la estación pierde la energía que la solventa y varios de ellos, al notar este fallo, corren para examinar el contador y eventualmente digitar los números manualmente. Sin embargo, al encontrarse dentro del

laboratorio digitando los números el temporizador programado por la máquina se activa e implosiona el laboratorio.

En el primer y segundo ejemplo los *affordances* pueden estar limitados solo a una serie de conductas, debido a que los artefactos entramados en el sistema ofrecen pocas posibilidades. Sin embargo, en el primer escenario los personajes asumen gran parte de estas tareas incluso al accionar los artefactos que están implicados en una serie de conductas. Además, el crédito epistémico y moral por estas acciones es asumido por los agentes porque siguen involucrando sentimientos debido al mérito personal, el rechazo o la satisfacción por haber desarrollado una tarea. También, a pesar de que en principio no conocen cómo funciona la computadora, ellos sí conocen parte del proceso que involucra la digitación de los números, el contador de tiempo y varias tareas más dentro del laboratorio como encender las luces, abrir las puertas, etc. En pocas palabras, los agentes aún conservan una fenomenología moral que les permite *saber actuar como* agentes con sentimientos.

En cuanto al segundo ejemplo, ocurre algo distinto. Los agentes comienzan a delegar gran parte de sus acciones cotidianas en el laboratorio a la máquina autónoma inteligente. Algo similar ocurre con las instituciones y los agentes de tránsito cuando delegan la función de regular la velocidad a los reductores incrustados en las carreteras. Pero en este segundo escenario, la máquina inteligente incluso desarrolla otras tareas en el laboratorio. Conforme transcurre el tiempo, los agentes experimentan un detrimento o una pérdida del crédito epistémico y moral porque parte del enriquecimiento de aspectos como el mérito personal involucra la aceptación de la satisfacción o el rechazo según ciertas conductas. En parte, al delegar totalmente la función de pulsar los números, la fenomenología moral que acompaña a esta conducta se pierde, lo que implica a su vez un detrimento. Por último, parte del determinismo radica en que la programación de esta máquina permanece oculta para los personajes, incluso en parte de su uso inconsciente (como ocurre con la transparencia), por lo que el agente pierde su capacidad crítica de determinar si realmente la máquina puede cumplir o no su función de forma adecuada.

De fondo lo que deseo señalar aquí es que las máquinas autónomas carecen de esta fenomenología moral porque es una ventaja evolutiva de los organismos que logran vivir según unos valores como, por ejemplo, el mérito personal. Esto no es algo que obedezca a un esencialismo genético por el que operemos los seres humanos. Más bien, la fenomenología moral es un conjunto de experiencias que logran algunos organismos en una interacción con artefactos, el entorno y otros agentes. Pero el peligro de estos contextos particulares es que los agentes e incluso algunos entornos podrían verse expuestos a determinismos artefactuales.

Nyholm y Smids (2016, 2020) señalan casos particulares en los que la automatización de máquinas puede crear contextos como el clásico dilema del tranvía. Por ejemplo, los vehículos autónomos, las estaciones de tren o el tráfico aéreo podrían crear entornos automatizados complejos que proponen también dilemas complejos. El mismo Nyholm (2018) reflexiona sobre la automatización y la dificultad de adjudicar responsabilidad sobre casos particulares que involucran máquinas y agentes humanos. Al respecto quisiera señalar que en los casos de determinismo artefactual la responsabilidad podría ser ambigua ya que, si se cumplen los criterios propuestos, la delegación de

varios aspectos de la conducta humana conduciría incluso a delegar la responsabilidad. Finalmente, de acuerdo con Nyholm (2024) y Rivera-Novoa (2024), la pérdida de mérito personal o cierto orgullo por haber realizado una tarea podría ser un indicio de un criterio para determinar cuándo hay o no detrimento en la fenomenología moral. Esto último podría ser motivo de exploración sobre mejoras morales en un artículo posterior.

CONCLUSIONES

El genocentrismo sostiene que el comportamiento humano obedece a una serie de propiedades vinculadas con la información contenida en los genes. En consecuencia, el comportamiento o la conducta humana estarían íntimamente ligados a aspectos genéticos y hereditarios. Contra esta visión, los avances en la biología y las teorías cognitivistas *4E* mostrarían que el entorno tiene un rol activo en cómo se configuran los organismos. Ahora bien, tampoco el entorno influye de forma unidireccional en cómo actúan los agentes, sino que la conducta se modela por un entramado entre el entorno, los agentes y los artefactos tecnológicos.

La tesis de la mente extendida sirve como marco explicativo para mostrar cómo los artefactos pueden involucrar procesos cognitivos que permiten desarrollar actividades. También esta misma tesis ofrecería algunas alternativas para pensar cómo mediante la manipulación o complementariedad de estos artefactos pueden vincularse agentes con herramientas tecnológicas.

El concepto de transparencia explica cómo el vínculo de los agentes con los artefactos puede conducir a escenarios en los que el agente desconozca el trasfondo del funcionamiento de las herramientas tecnológicas. Cuando la transparencia es «en el uso», el énfasis del agente está en la atención que pone a la tarea que desea realizar según ciertos objetivos y no en el elemento material. En consecuencia, parte del *black-boxing* de los artefactos permanece oculto al agente.

El determinismo artefactual es un caso particular en el que un agente dentro de un sistema genuino de mente extendida se ve determinado a ciertas conductas debido a las *affordances* de máquinas autónomas. Los criterios de delegación de tareas, el detrimento del crédito moral y la transparencia mostrarían cómo en la interacción con ciertos artefactos tecnológicos autónomos como las IA, algunos aspectos como la responsabilidad o el mérito personal pueden ser ambiguos y, por lo tanto, ofrecen dilemas complejos sobre la conducta agencial. La fenomenología moral podría proporcionar criterios para establecer cuándo hay o no detrimento o mejoras morales.

El determinismo artefactual argumentado muestra que hay casos en los que artefactos tecnológicos autónomos, como las IA, conducen a los agentes biológicos a un rol pasivo. En este sentido, a pesar de que la tesis de la mente extendida sea un marco explicativo para argumentar contra el genocentrismo, también mostraría que artefactos como las IA autónomas determinarían la conducta humana planteando escenarios complejos como la pérdida del crédito moral y la autonomía.

AGRADECIMIENTOS

Agradezco a Carolina Marulanda y a Leandro Giri, quienes con sus observaciones y comentarios dieron lugar a esta versión del artículo. Asimismo, agradezco al profesor Ángel Rivera Novoa, quien me invitó a participar en este proyecto de investigación. También a los evaluadores de este artículo, pues con sus comentarios contribuyeron al texto publicado.

CONFLICTOS DE INTERÉS

Este artículo no tiene posible conflicto de interés de orden financiero, profesional o personal, que pueda influir de forma inapropiada en los resultados obtenidos o las interpretaciones propuestas en el texto aquí presentado.

REFERENCIAS

Abrams, J. J., y Lindelof, D. (2004). *Lost* [serie]. American Broadcasting Company.

Andrada, G., Clowes, R. W., y Smart, P. R. (2023). Varieties of transparency: Exploring agency within AI systems. *AI & SOCIETY*, *38*(4), 1321-1331. https://doi.org/10.1007/s00146-021-01326-6

Anscombe, G. E. M. (1991). *Intención*. Ediciones Paidós.

Clark, A., y Chalmers, D. (1998). The Extended Mind. *Analysis*, *58*(1), 7-19. https://doi.org/10.1093/analys/58.1.7

Crick, F. H. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology*, *12*, 138-163.

Darwin, C. (1859). *On the Origin of Species: Facsimile of the First Edition*. John Murray, Albemarle Street.

Darwin, C. (1872). T*he expression of emotions in animals and man*. Murray.

Davidson, D. (1963). Actions, Reasons, and Causes. *The Journal of Philosophy*, *60*(23), 685. https://doi.org/10.2307/2023177

Davidson, D. (1971). I. Agency. En R. W. Binkley, R. N. Bronaugh, y A. Marras (eds.), *Agent, Action, and Reason* (pp. 1-37). University of Toronto Press. https://doi.org/10.3138/9781442656963-002

Dawkins, R. (2006). *The selfish gene* (30th anniversary ed). Oxford University Press.

Di Paolo, E. A., Buhrmann, T., y Barandiaran, X. E. (2017). *Sensorimotor life: An enactive proposal*. Oxford University Press.

Fasoli, M. (2018). Substitutive, Complementary and Constitutive Cognitive Artifacts: Developing an Interaction-Centered Approach. *Review of Philosophy and Psychology*, *9*(3), 671-687. https://doi.org/10.1007/s13164-017-0363-2

Fuchs, T. (2017). *Ecology of the Brain: The phenomenology and biology of the embodied mind*. Oxford University Press. https://doi.org/10.1093/med/9780199646883.001.0001

Gibson, J. J. (2015). *The Ecological Approach to Visual Perception*. Psychology Press.

Heidegger, M. (2016). *Ser y tiempo* (J. E. Rivera, trad.) (2.ª ed.). Trotta.

Hume, D. (1990). *An Enquiry Concerning the Principles of Morals* (J. B. Schneewind, ed.). Hackett.

Kirchhoff, M. D., y Kiverstein, J. (2019). *Extended Consciousness and Predictive Processing: A Third-Wave View*. Routledge.

Kiverstein, J. (2018). Extended Cognition. En A. Newen, L. De Bruin, y S. Gallagher (eds.), *The Oxford Handbook of 4E Cognition* (pp. 18-40). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.013.2

Laland, K. N., Odling-Smee, J., y Feldman, M. W. (2000). Niche construction, biological evolution, and cultural change. *Behavioral and Brain Sciences*, *23*(1), 131-146. https://doi.org/10.1017/S0140525X00002417

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press.

Lewontin, R. C., Rose, S., y Kamin, L. J. (1984). *Not in Our Genes: Biology, Ideology, and Human Nature*. Pantheon Books.

Lippman, A. (1993). Prenatal Genetic Testing and Geneticization: Mother Matters for All. *Fetal Diagnosis and Therapy*, *8*(1), 175-188. https://doi.org/10.1159/000263886

Lykken, D. (2000). *Happiness: The Nature and Nurture of Joy and Contentment*. St. Martin's Griffin.

Menary, R. (2007). *Cognitive Integration: Mind and Cognition Unbounded*. Palgrave Macmillan.

Menary, R. (ed.). (2010). Cognitive Integration and the Extended Mind. En *The Extended Mind* (pp. 227-244). The MIT press.

Monterroza-Ríos, Á. D. (2018). *La naturaleza heterogénea de los artefactos. Un análisis ontológico*. Editorial ITM. http://hdl.handle.net/20.500.12622/1824

Monterroza-Rios, A. D. (2023). Una Concepción Enactiva de Cultura: Enculturación como Acople Dinámico entre Seres Humanos y sus Entornos de Cultura Material. *Principia: an international journal of epistemology*, *27*(2), 215-234. https://doi.org/10.5007/1808-1711.2023.e85209

Monterroza-Rios, A. D., y Gutiérrez-Aguilar, C. M. (2022). Enactivism and Material Culture: How Enactivism Could Redefine Enculturation Processes. *Philosophies*, *7*(4), 75. https://doi.org/10.3390/philosophies7040075

Nelkin, D., y Lindee, M. S. (1995). *The DNA Mystique: The Gene as a Cultural Icon*. Freeman.

Newen, A., Gallagher, S., y De Bruin, L. (2018). 4E Cognition: Historical Roots, Key Concepts, and Central Issues. En *The Oxford Handbook of 4E Cognition* (pp. 2-16). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.013.1

Nyholm, S. (2018). Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility-Loci. *Science and Engineering Ethics*, *24*(4), 1201-1219. https://doi.org/10.1007/s11948-017-9943-x

Nyholm, S. (2024). Artificial Intelligence and Human Enhancement: Can AI Technologies Make Us More (Artificially) Intelligent? *Cambridge Quarterly of Healthcare Ethics*, *33*(1), 76-88. https://doi.org/10.1017/S0963180123000464

Nyholm, S., y Smids, J. (2016). The Ethics of Accident-Algorithms for Self-Driving Cars: An Applied Trolley Problem? *Ethical Theory and Moral Practice*, *19*(5), 1275-1289. https://doi.org/10.1007/s10677-016-9745-2

Nyholm, S., y Smids, J. (2020). Automated cars meet human drivers: Responsible human-robot coordination and the ethics of mixed traffic. *Ethics and Information Technology*, *22*(4), 335-344. https://doi.org/10.1007/s10676-018-9445-9

Parselis, M. (2018). *Dar sentido a la técnica: ¿pueden ser honestas las tecnologías?* Catarata.

Persson, I., y Savulescu, J. (2008). The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity. *Journal of Applied Philosophy*, *25*(3), 162-177. https://doi.org/10.1111/j.1468-5930.2008.00410.x

Rivera-Novoa, A. (2024). La tesis de la mente extendida y el ideal transhumanista de mejoramiento cognitivo. *Trilogía Ciencia Tecnología Sociedad*, *16*(33), e3142. https://doi.org/10.22430/21457778.3142

Rosoff, P. M., y Rosenberg, A. (2006). How Darwinian reductionism refutes genetic determinism. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *37*(1), 122-135. https://doi.org/10.1016/j.shpsc.2005.12.005

Shapiro, L., y Spaulding, S. (2024). Embodied Cognition. En E. N. Zalta, y U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab. https://plato.stanford.edu/archives/fall2024/entries/embodied-cognition/

Sharon, T. (2013). *Human Nature in an Age of Biotechnology: The Case for Mediated Posthumanism*. Springer Berlin Heidelberg.

Sutton, J. (2010). Exograms and Interdisciplinarity: History, the Extended Mind, and the Civilizing Process. En R. Menary (ed.), *The Extended Mind* (pp. 189-225). MIT Press.

Tomasello, M. (2003). *The Cultural Origins of Human Cognition*. Harvard University Press.

Turilli, M. (2007). Ethical protocols design. *Ethics and Information Technology*, *9*(1), 49-62. https://doi.org/10.1007/s10676-006-9128-9

Turilli, M., y Floridi, L. (2009). The ethics of information transparency. *Ethics and Information Technology*, *11*(2), 105-112. https://doi.org/10.1007/s10676-009-9187-9

Turkheimer, E. (2000). Three Laws of Behavior Genetics and What They Mean. *Current Directions in Psychological Science*, *9*(5), 160-164. https://doi.org/10.1111/1467-8721.00084

Varela, F. J., Thompson, E., y Rosch, E. (1993). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.

Verbeek, P.-P. (2005). *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. Pennsylvania State University Press.

Wilson, E. O. (1975). *Sociobiology: The New Synthesis*. Belknap Press of Harvard University Press.