Sesgos cognitivos, pseudociencias e inteligencia artificial

Discurso de Recepción de la
Académica de Número
Excma. Sra. Dª. Helena Matute Greño

y

contestación del
Excmo. Sr. D. Enrique Echeburúa Odriozola

Sesión Pública del 6 de noviembre de 2024
Madrid

Sesgos cognitivos, pseudociencias e inteligencia artificial

Excma. Sra. Dª. Helena Matute Greño

Agradecimientos

Excelentísimo Sr. Presidente de la Academia de Psicología de España, excelentísimos miembros de la Academia, amigas y amigos.

Es para mí un gran honor estar hoy aquí con todos ustedes. Ser elegida como miembro de esta ilustre Academia es algo en lo que no había siquiera soñado, y debo dar las gracias a muchas personas por haberlo hecho posible.

Quiero dar las gracias en primer lugar a los tres académicos que presentaron mi candidatura, los profesores Enrique Echeburúa, Pío Tudela, y Jaime Vila. Me ha hecho mucha ilusión que fueran ellos quienes presentaron mi candidatura porque siempre han sido tres grandes referentes para mí. Además, debo agradecer muy especialmente al profesor Echeburúa, a quien siempre he admirado por su trabajo pionero y comprometido en el País Vasco, su generosidad por haber realizado la laudatio, y por aceptar también la encomienda de responder hoy a mi discurso, en representación de la Academia.

Muchas gracias también al presidente de la Academia, el profesor Helio Carpintero, que junto al profesor Echeburúa me ha ayudado y guiado en todo el proceso. Y gracias también, de corazón, a todos los académicos que han confiado en mí y han apoyado con su voto mi nombramiento.

Quiero dar las gracias también a todo mi equipo de investigación, mi querido Labpsico. Jóvenes brillantes, muchos de ellos hoy profesores e investigadores diseminados por universidades de todo el mundo, algunos trabajando todavía conmigo en su tesis doctoral en nuestro Laboratorio de Psicología Experimental, un hervidero de ideas, discusiones e ilusiones, donde da gusto trabajar. Creo que lo mejor de mi trabajo, y el gran truco para estar hoy aquí con todos ustedes recibiendo este reconocimiento al trabajo realizado, no es otro que haber sabido siempre rodearme de los mejores. Estoy muy orgullosa de todos vosotros, Labpsiqueros. Me siento muy orgullosa si he podido contribuir algo en vuestra formación, y me siento muy afortunada por todo lo que aprendo de vosotros, por todas las discusiones apasionantes que mantenemos y todas las ilusiones que compartimos.

Gracias también a mi familia y amigos, por estar siempre ahí, en los momentos buenos y en los malos, dándome siempre el apoyo necesario para seguir adelante. Gracias por vuestro apoyo, y gracias por estar siempre conmigo cuando ha hecho falta.

Y a Juan Ángel, mi querido compañero de vida. Son muchos los momentos que te he robado, y mucho lo que tú me has dado: paciencia, amor, alegría, fuerza, siempre he contado con tu apoyo. También con tus grandes y estimulantes discusiones durante nuestras caminatas por el monte y por la playa. Gracias por estar conmigo. Nada de esto sería posible sin ti.

Y a la Academia de Psicología de España: ¡muchas gracias! No sé si merezco este nombramiento, pero lo tomo como una gran responsabilidad y espero llegar a merecerlo algún día.

Gracias a todos, de corazón.

1.- Introducción

Los que hacemos investigación experimental sobre los procesos psicológicos básicos solemos recibir a menudo preguntas (o miradas) del tipo de, "¿Y eso para qué sirve?"

Y a esto contestamos siempre explicando que "la investigación básica es lo más práctico (y lo más bonito) que hay", porque, además, "no te centras en ninguna aplicación concreta, sino en el avance del conocimiento". Y eso significa, que al final podrás aplicarlo a cualquier tema que te interese… Incluso puedes inventarte aplicaciones nuevas. Y eso, es muy bonito.

De esto último es de lo que quiero hablarles hoy. De una serie de áreas nuevas en las que estamos trabajando estos últimos años, que creo que tienen gran interés para la psicología y además tienen vínculos con otras disciplinas, como la economía, el derecho, la informática, y la sanidad.

A primera vista cualquiera nos diría que no tienen nada que ver unas con otras, pero estoy segura de que esta audiencia habrá ya adivinado que es precisamente la psicología humana la que nos proporciona un nexo común cuando hablamos de áreas de actualidad, como la inteligencia artificial, el cambio climático, la salud pública, las desigualdades sociales… Son todos ellos campos donde los psicólogos tenemos mucho que aportar, pues todos ellos dependen de comportamientos, actitudes, creencias, emociones, y también de sesgos, prejuicios, y errores humanos.

Estoy convencida, además, de que los psicólogos debemos trabajar en estos nuevos campos, y que cuanto antes lo hagamos, mucho mejor, porque son muchos los profesionales de otras disciplinas los que, a falta de psicólogos que se ocupen muchas veces de abordar estos problemas, empiezan a hablar también de mente y conducta humana, de emociones y razonamientos, de aprendizaje y sesgos cognitivos, de toma de decisiones, de fobias y filias, llenando un hueco que los psicólogos no siempre estamos llenando, cuando se trata, sin embargo, de conceptos que están en la formación básica de todo psicólogo, y en el centro de la investigación básica en psicología.

Y en esto creo que esta Academia, si se me permite, puede tener probablemente un gran papel, no solo a la hora de abrir campos nuevos para nuestra disciplina, sino también a la hora de animar a los nuevos psicólogos a involucrarse en ellos y a hacer de la psicología la ciencia relevante para la humanidad que puede y debe llegar a ser desde ámbitos muy diversos.

Preparando este discurso de ingreso leí el discurso que la Académica Mª Paz García-Vera pronunció en el año 2020 (véase García-Vera, 2023), donde ponía de manifiesto cómo los psicólogos nos hemos ocupado tradicionalmente de la psicología clínica dejando de lado muchas otras áreas en las que podemos y debemos ser relevantes. Ella pone el ejemplo de la pandemia de COVID-19, y me parece muy pertinente.

¿Cómo puede ser que durante muchos meses las únicas vacunas que existieron para frenar la pandemia fueron comportamentales (distancia social, mascarilla, higiene de manos...), y sin embargo los psicólogos no fuimos una pieza clave en el desarrollo de estrategias de control de la pandemia, ni en la configuración de los mensajes que se transmitían la sociedad, o en cómo se intentaba motivar (o no) los comportamientos críticos?

No entraré en detalles, pues ya lo trató magistralmente García-Vera (2023) en su propio discurso, y además yo no he investigado directamente el tema de las pandemias; pero me interesa mucho el ejemplo pues algo no muy diferente es lo que creo que nos está pasando con otros muchísimos temas que sí conozco en mayor detalle y en los que creo que deberíamos ser mucho más relevantes de lo que estamos siendo.

Desde esta perspectiva quisiera comentar a continuación dos de las líneas en las que estamos trabajando en mi grupo de investigación. En ambos casos, partiendo del estudio experimental de algunos de los procesos psicológicos básicos (especialmente el aprendizaje, los sesgos cognitivos, y la toma de decisiones), estamos explorando una serie de cuestiones que creo que son de máxima relevancia en el contexto actual. Me centraré en dos de ellas, que aparentemente poco o nada tienen en común: Las pseudociencias y la inteligencia artificial (IA), o más concretamente, las pseudociencias y la relación de las personas con la IA, y qué papel juegan, en ambos casos, los sesgos cognitivos.

Los sesgos cognitivos son errores sistemáticos y predecibles en nuestro procesamiento de la información (Kahneman, 2011). Tanto las pseudociencias como la inteligencia artificial pueden hacer uso de ellos y explotar los sesgos humanos, y creo que los psicólogos estamos mejor preparados que otros colectivos para dar respuesta a este problema. Ambos temas constituyen, además, fuentes inagotables de experimentos que podemos realizar sobre el aprendizaje (humano, animal, y ahora también, artificial), la toma de decisiones, el pensamiento mágico (ante la pseudociencia y ante la IA), y

por supuesto, los sesgos cognitivos (en humanos, y en IAs). Y ambos temas pseudociencias e IA, son apasionantes. Empezaré por las pseudociencias.

2.- Pseudociencias

El diccionario de la Real Academia de la Lengua Española define el adjetivo pseudocientífico como "falsamente científico". La palabra pseudociencia como tal no aparece en el diccionario, pero podemos definirla como toda práctica o creencia que se presenta como científica, pero que carece de evidencia que la sustente.

Un gran ejemplo es el bálsamo de Fierabrás, que tan bien caricaturizó Cervantes. Según Don Quijote, después de la batalla, el bálsamo sería capaz de curar cualquier herida, por muy grave que ésta fuera. Pero cuando el pobre Sancho probó el bálsamo, le hizo tanto daño que creyó morir, a lo que Don Quijote contestó negando la evidencia, y explicándole que el problema no era el bálsamo, sino que a él no le podía sentar bien puesto que no era caballero.

Bromas cervantinas aparte, las pseudociencias son un problema serio. Hay gente que muere por no ir al hospital, por no tratar a tiempo una enfermedad, debido a que están convencidos de que podrán curarse con remedios que no tienen ninguna efectividad (Freckelton, 2012). Hay también gente que gasta mucho dinero en remedios que no sirven para nada; hay gente que inventa conspiraciones, y fomenta la desinformación, con tal de poder difundir determinadas teorías pseudocientíficas que, al igual que Don Quijote, creen firmemente que son correctas a pesar de que no sean avaladas por la evidencia. Simplemente creen que un día alguien demostrará que son correctas, y que por tanto es lícito, según ellos, fomentar su uso.

Las pseudociencias pueden afectar a la salud (FECYT 2020; 2022; OMC, s/f), como estamos viendo, pero también a la educación (donde también se aplican a diario técnicas no probadas, Double et al., 2020; Ferrero et al., 2016; 2020), a la economía y las finanzas (Ferreira, 2015), o a la forma de cultivar las semillas de tomate (Mulet, 2015), por poner una serie de ejemplos que afectan a áreas muy diversas de la vida de las personas.

También en psicología tenemos, por supuesto, nuestras propias pseudociencias, de las que no acabamos de desprendernos. Es cierto que somos una de las ciencias que más está luchando en el momento actual por mejorar sus

métodos, y la reproducibilidad de sus investigaciones, con grandes proyectos en los que se involucran muchos laboratorios (e.g., Klein et al., 2014; Open Science Collaboration, 2015). Pese a ello, y volviendo a las pseudociencias, tenemos que reconocer que aún hoy en día, en psicología contamos con numerosos ejemplos de terapias no avaladas por la evidencia científica, y que sin embargo se siguen aplicando a diario, bajo el argumento de que parece que funcionan, y que quizá un día se demuestre que lo hacen. Como dijo el profesor Enrique Echeburúa en su discurso de ingreso en la Academia en 2022:

"No es exagerado afirmar que las terapias psicológicas no validadas empíricamente se utilizan en la clínica con más frecuencia que los tratamientos basados en pruebas y que, por tanto, hay un desfase entre lo que se sabe y lo que se hace… De hecho, los clínicos (y no solo por la presión asistencial) se muestran muy reticentes a incorporar novedades clínicas validadas empíricamente cuando esto supone modificar modelos teóricos arraigados (la resistencia al cambio) o líneas de actuación y formas de abordar los problemas que tienen sobreaprendidas" (véase Echeburúa, 2023, pág. 236)

Esto es grave. Y es precisamente la base de las pseudociencias, el dar por hecho que mi experiencia, y mi intuición personal, tienen más valor que años de investigaciones rigurosas y contrastadas para saber cuándo algo funciona y cuándo no lo hace.

La pseudociencia no se da solo en solo en psicología (ni solo en la clínica, por supuesto), se da en todos los ámbitos, pero somos los psicólogos los que deberíamos estar más interesados en erradicarla de nuestra propia disciplina, y los que estamos además más capacitados para ayudar a erradicarla también de otras disciplinas, ya que se trata de un problema de creencias, actitudes y conductas. Es un problema psicológico. Nos corresponde abordarlo.

Por tanto, la pregunta es: ¿Qué podemos aportar los psicólogos para reducir este problema en nuestra ciencia y en otras?

Mi primera respuesta es: investigación. Existe, por una parte, mucha investigación básica disponible en psicología sobre creencias, actitudes, conducta, aprendizaje, toma de decisiones, sesgos cognitivos, que puede ayudar a encauzar el problema y plantear, por otra parte, hipótesis específicas y experimentos más aplicados sobre pseudociencias, que nos ayuden a entender mejor su funcionamiento y a desarrollar estrategias de prevención y de afrontamiento.

Y mi segunda respuesta es: divulgación de la investigación a la sociedad, es decir, educación ciudadana sobre el funcionamiento de la mente, el cerebro, la conducta humana, sobre los principios de aprendizaje y conducta, sobre procesos psicológicos básicos, emociones, personalidad, sobre las pseudociencias. Y, muy importante, sobre metodología científica, porque, al fin y al cabo, el método científico es una herramienta psicológica: es la mejor herramienta que ha inventado la humanidad para reducir el impacto de los sesgos cognitivos y por tanto también para reducir la pseudociencia (Lilienfeld et al., 2011).

De todo esto les hablaré a continuación.

2.1. – Experimentos sobre la ilusión (o sesgo) de causalidad

Algo que me parece importante plantearnos como psicólogos es por qué hay personas que prefieren utilizar un tratamiento no avalado por la ciencia frente a otro científicamente contrastado. O por qué hay personas que prefieren utilizar en su docencia técnicas novedosas que nadie ha demostrado que sean efectivas. Las pseudociencias, como ya hemos comentado, afectan a la psicología y también a la educación, la medicina, la enfermería, la economía, el deporte, la empresa… y en general a todas las disciplinas en las que las creencias y las decisiones humanas son importantes.

Existen muchos factores que pueden hacer que algunas personas confíen más en tratamientos y técnicas no probados (e.g., FECYT, 2020; 2022; Piejka y Okruszek, 2020; Lobera y Rogero-García, 2021; Rodríguez-Ferreiro y Barberia, 2021), pero uno muy importante en el que quisiera detenerme es que esas personas creen que a ellas les funciona (o les va a funcionar), independientemente de que existan o no pruebas científicas que lo validen (Segovia y Sanz-Barbero, 2022).

Por tanto, entre otras muchas cosas, debemos averiguar qué les hace creer que algo les funciona, cuando no funciona; o, dicho en otras palabras, qué hace que se desarrolle esa ilusión de que X es causa de Y cuando en realidad no lo es. Esto es lo que llamamos ilusión de causalidad, y es uno de los principales sesgos cognitivos que están en la base de las pseudociencias (Matute et al., 2011; Torres et al., 2020).

Llevamos muchos años en mi grupo de investigación estudiando las ilusiones de causalidad, las variables que influyen, en qué situaciones se produce mejor aprendizaje de causa-efecto, en cuáles se da más ilusión (de que X es causa de Y cuando no lo es), y también, por supuesto, de qué manera podríamos disminuir esa ilusión causal.

La idea es que si averiguamos qué variables influyen y cómo desarrollan las personas esa ilusión de causa-efecto, podremos desarrollar estrategias para reducirla, y podríamos aplicarlas en cualquier campo que sea de nuestro interés, incluida la psicología.

Para simplificar los experimentos que realizamos, y para evitar problemas éticos y la influencia de problemas adicionales (como por ejemplo el efecto placebo o los efectos relacionados con el conocimiento previo y posibles prejuicios), todos los experimentos los realizamos por ordenador, con fichas médicas ficticias de pacientes ficticios, que siguen o no unos tratamientos, también ficticios, para dolencias ficticias. Son experimentos en los que estudiamos qué hace que los participantes perciban que existe una relación causal entre por ejemplo una pseudoterapia a la que se están sometiendo una serie de pacientes, y el resultado deseado (la mejoría observada en un porcentaje elevado de los pacientes). En realidad, el ordenador está programado para que se curen muchos pacientes, pero siempre con la misma probabilidad tanto si siguen el tratamiento como si no lo siguen. Es decir, se trata de una pseudoterapia que no tiene ningún efecto. La cuestión es en qué situaciones las personas desarrollan más ilusión de que la pseudoterapia está funcionando y cómo podríamos combatir este problema (véase Matute et al., 2015; 2019 para una revisión).

Imaginemos que una persona toma unas píldoras alternativas que le ha recomendado un amigo para tratar una jaqueca. El día siguiente se encuentra mejor, e inmediatamente asocia su mejoría con la píldora que tomó la noche anterior. En este ejemplo no hay ningún tipo de relación causal comprobada, pues lo único que observamos es una contigüidad temporal entre los dos eventos. Pero la persona puede fácilmente sacar la conclusión de que la relación es causal y que el remedio funciona. De hecho, la contigüidad temporal es una de las claves que más utilizamos todos para poder inferir causalidad en nuestro día a día (Hume, 1741/1978; Shanks et al., 1989; Wasserman y Neunaber, 1986), ya que saber con certeza si hay o no relación causal es difícil. Intuir la existencia de una relación causal a través

de la contigüidad temporal es relativamente sencillo, y a menudo se trata de un heurístico, o atajo cognitivo, bastante acertado; pero también nos puede provocar este tipo de sesgos de casualidad, como estamos viendo.

Para medir estas ilusiones en estos experimentos, después de observar una serie de fichas de pacientes ficticios que siguen (o no) el tratamiento pseudocientífico (ficticio), y se curan (o no), preguntamos al participante sobre el grado en que cree que el tratamiento es la causa de las curaciones observadas. Dado que podemos manipular la secuencia de acontecimientos para que se cure el mismo porcentaje de pacientes tanto si siguen el tratamiento como si no lo siguen, sabemos con certeza que la causalidad real entre el tratamiento y la curación es cero.

Sin embargo, los participantes suelen dar respuestas superiores a cero. Es decir, suelen desarrollar ilusión casual. Y lo interesante en estos casos es que el grado en que se desarrolla esta ilusión causal depende de una serie de variables que podemos manipular en los experimentos y que conocemos cada vez mejor cómo influyen, lo cual nos sirve para conocer mejor el funcionamiento de las pseudociencias, en qué casos la gente creerá que productos que son auténticas estafas, tienen un efecto beneficioso, y qué podemos hacer para disminuir este problema.

No quiero detenerme a describir los detalles de los resultados de estos experimentos, los trabajos están publicados y pueden consultarse (para una revisión véase Matute et al., 2015; 2019). Quisiera destacar únicamente un pequeño resumen de algunos de los resultados más robustos, sobre las variables que más claramente influyen en el desarrollo de este sesgo.

La ilusión de causa-efecto aumenta cuando la probabilidad de ocurrencia del resultado deseado (las curaciones), o la probabilidad de ocurrencia de la causa (la frecuencia con que se administra el tratamiento), o ambas, son elevadas (Allan et al., 2005; Blanco et al., 2013; Chow et al., 2019; Hannah y Beneteau, 2009; Perales et al., 2005), de modo que puedan darse muchas coincidencias entre ambos eventos, y por lo tanto, se fortalezca así la asociación entre ambos. De hecho, las pseudoterapias en la vida real suelen aplicarse con frecuencia en situaciones en las que existe un alto porcentaje de remisiones espontáneas de la enfermedad (dolor de cabeza, dolor de espalda, ansiedad, depresión, malestar general…), y además, muchos tratamientos pseudocientíficos recomiendan al paciente tomar el medicamento con

frecuencia (con idea de que haya muchas coincidencias entre la causa y el efecto potenciales, y puedan quedar de esa forma fuertemente asociados, dando lugar a la ilusión de que el tratamiento funciona).

Además, si estamos diciendo que en general la probabilidad de la causa y del efecto (en definitiva, las coincidencias) deben ser elevadas para que se desarrollen estas ilusiones, podemos predecir también una serie de efectos muy interesantes. Así, en situaciones en las que el resultado deseado ocurre con frecuencia (por ejemplo, dolencias que remiten espontáneamente) podemos predecir que se dará más ilusión causal si el participante se involucra personalmente en el tratamiento (ya que en esos casos normalmente aplicará la causa con mayor frecuencia, lo que dará lugar a más coincidencias y por tanto a un mayor sesgo de causalidad) (Yarritu et al., 2014). Algo parecido ocurre también cuando, por ejemplo, los participantes tienen más recursos para comprar el medicamento (ya que en esos casos lo administran con más frecuencia), lo que una vez más da lugar a un mayor grado de ilusión causal (Vinas et al., 2023). Sin ser la única explicación posible, este efecto podría explicar también, al menos en parte, el avance actual de las pseudociencias en los países ricos.

No pretendo realizar aquí una revisión exhaustiva de todas las variables que influyen en la ilusión causal, ni mucho menos de todas las que influyen en el desarrollo de las pseudociencias, que resultan, así mismo, influenciadas no solo por este sesgo de ilusión causal que estoy describiendo sino también por muchos otros, como por ejemplo la tendencia a saltar a la conclusión (Moreno-Fernández et al., 2021; Rodríguez-Ferreiro y Barberia, 2021). Sirvan los ejemplos mencionados para mostrar que es mucho lo que podemos aportar los psicólogos, no solo a la reducción de las pseudociencias en nuestra propia ciencia, sino también en otras, y en general, para ayudar a los ciudadanos a protegerse contra este tipo de sesgos.

2.2. – Divulgación psicológica y educación ciudadana

Me hace mucha ilusión, por tanto, y para finalizar esta primera parte dedicada a la investigación sobre las pseudociencias, contarles que hemos realizado recientemente una colaboración con la Fundación Española para la Ciencia y la Tecnología (FECYT), que nos ha permitido llevar estas investigaciones a más de 40 colegios de toda España, en los que más de 2000 adolescentes han

podido beneficiarse del programa de intervención que hemos desarrollado para reducir la ilusión causal, y hemos podido comprobar además, tras un seguimiento a seis meses, que los efectos son duraderos al menos a medio plazo (Martínez et al., 2024).

La intervención incluía una primera fase de inducción del sesgo, en la que hicimos creer a los adolescentes que un anillo milagroso que les presentamos con mucha jerga científica, servía para aumentar sus capacidades físicas y cognitivas. Tuvieron oportunidad de probarlo, pero las condiciones de control siempre fueron malas o inexistentes, de modo que creyeron en las capacidades del anillo y la historia que les habíamos contado. En una segunda fase, les mostramos cómo les habíamos engañado y qué deberían saber, preguntar, probar, cuando quieran saber en el futuro si algo funciona o no, para que no les vuelvan a engañar tan fácilmente; es decir, la segunda parte consistía en un taller sencillo de metodología científica y control de variables ante una audiencia de adolescentes que estaban, ahora sí, muy receptivos a la información que queríamos trasmitirles sobre cómo protegerse de ciertas ilusiones cognitivas y engaños pseudocientíficos.

Para medir posteriormente si había sido eficaz la intervención, utilizamos un experimento estándar de ilusión causal, vía Internet, y similar a los que he descrito anteriormente. Los resultados mostraron que los estudiantes que habían realizado la intervención fueron capaces de buscar las pruebas necesarias y reducir significativamente sus ilusiones causales en comparación con los estudiantes que no la habían realizado. Además, seis meses después seguían mostrando menos ilusión causal aquellos que habían recibido la intervención, en comparación con que los que no la habían recibido.

Teniendo en cuenta que se trataba de una intervención muy sencilla de poco más de una hora de duración en la que simplemente mostrábamos las bases del pensamiento científico para minimizar los propios sesgos, esto nos da esperanzas de que se pueda seguir aplicando la intervención a través del sistema educativo, para poder dotar a todos los ciudadanos de unas herramientas básicas que les ayuden a detectar los sesgos, y enfrentarse de forma más crítica y empoderada a las situaciones relacionadas con la pseudociencia con las que se encuentren a lo largo de su vida.

Estamos especialmente orgullosos de este proyecto en el que, gracias a la colaboración con FECYT, hemos podido comprobar cómo los resultados de los experimentos de laboratorio en los que llevábamos muchos años trabajando, pueden resultar útiles y extenderse a través del sistema educativo a programas de intervención a nivel nacional con resultados contrastables muy positivos. El trabajo es sencillo, pero no es muy habitual, créanme, para los que hacemos investigación de laboratorio, trabajar con muestras tan grandes ni poder realizar seguimientos a seis meses, ni trabajar en contextos reales, como es el caso de los colegios, en regiones tan diversas, y donde estudiantes reales se benefician directamente de nuestras investigaciones.

+ + + + + + + + +

Pero no quisiera cantar victoria tan pronto. Tal y como he comentado al inicio de mi intervención, existen en este momento muchas situaciones relevantes que inducen también con frecuencia el pensamiento mágico y una serie de reacciones e ilusiones cognitivas no muy diferentes de las que fomenta la pseudociencia. Me refiero a la otra gran área de investigación en la que les comenté que estamos trabajando y que está resultando también ser un auténtico filón para el desarrollo de experimentos sobre aprendizaje, sesgos cognitivos, errores en la toma de decisiones, y pensamiento mágico en contextos de máxima relevancia social. Me refiero a la inteligencia artificial, y especialmente a la relación de las personas con la IA.

3.- Inteligencia Artificial (IA)

A menudo nos han criticado a los psicólogos, que trabajamos con una caja negra en la que no podemos conocer lo que hay dentro. Introducimos unos estímulos (input), observamos la respuesta (output), pero la caja negra permanece fuera de nuestro control, y de nuestra comprensión. En el mejor de los casos, podemos inferir el contenido de la caja y su funcionamiento, y partir de ahí, hacer nuevas predicciones sobre lo que saldrá de la caja si introducimos tales inputs, predicciones que podemos poner a prueba haciendo nuevos experimentos y observando nuevamente las respuestas.

Además de hacer los experimentos, empezamos también los psicólogos hace muchos años a simular por ordenador lo que pensamos que había dentro de la

caja. El supuesto es que si tenemos una buena teoría sobre el funcionamiento de la caja negra (es decir, sobre el funcionamiento de la mente humana, o, siendo un poco más realistas, una buena teoría sobre el funcionamiento de al menos alguno de los procesos cognitivos, como por ejemplo, el aprendizaje), podríamos simular esa teoría en un programa de ordenador, y de esta forma podríamos hacer que el ordenador aprendiera y se comportara de manera similar a como predice nuestra teoría.

Por ejemplo, algunas de las teorías más populares del aprendizaje asociativo, como la de Rescorla y Wagner (1972), podían escribirse fácilmente en lenguaje de ordenador y podíamos poner a prueba sus predicciones observando si la máquina aprendía como lo hacían los participantes humanos y animales de nuestros experimentos.

Además, las teorías del aprendizaje asociativo eran especialmente interesantes en este contexto, porque partían del aprendizaje animal, pero estaban también siendo aplicadas con bastante éxito ya en los años 80 al estudio de aprendizajes complejos, como por ejemplo, el aprendizaje humano de relaciones causales y predictivas (Shanks, 2007), y podían explicar bastante bien, no todas, pero sí muchas de las situaciones en las que el aprendizaje se desviaba de la norma, como es el caso de muchos de los resultados básicos que hemos comentado en el apartado anterior donde las personas sobreestiman las relaciones de causalidad, desarrollando ilusiones de causa-efecto (Matute et al., 2015).

Además, tal y como ya abordó el profesor Pío Tudela en su discurso de ingreso en la Academia en 2021 (véase Tudela, 2023), estas simulaciones de modelos asociativos del aprendizaje no eran muy diferentes a los sistemas conexionistas de aprendizaje artificial que se estaban desarrollando también en los años 80 (Rumelhart y McClelland, 1986), y que son la base de muchos de los éxitos que se están logrando en la actualidad en el campo de la inteligencia artificial.

Y así, poco a poco, nos encontramos en la actualidad con nuevas cajas negras. Nuevos sistemas de inteligencia artificial, con una enorme capacidad de aprendizaje, con gran capacidad de adaptación y flexibilidad, pero por tanto también con gran capacidad de desarrollar sesgos y errores, como nos ocurre a las personas. Y son, una vez más, cajas negras.

Una vez más, no sabemos lo que hay dentro de la caja. Lo que tenemos es un algoritmo opaco, programado a menudo por una empresa tecnológica, con

unos objetivos y unos criterios que desconocemos, y además ese algoritmo va aprendiendo, y no tenemos forma de saber qué ha aprendido. No tenemos forma de saber de qué forma ni con qué criterios toma las decisiones que le pedimos que tome, cuando, por ejemplo, un banco utiliza un algoritmo de IA para decidir a quién se le concede o no una hipoteca, o cuando un juzgado lo utiliza para predecir qué reclusos tienen mayor probabilidad de reincidencia (O'Neil, 2016).

Por ello, y a pesar de las indudables ventajas y oportunidades que puede traer esta tecnología para la humanidad, se habla cada vez más de la necesidad de encauzarla adecuadamente, de la necesidad de auditar estos algoritmos que, cada vez más, gobiernan nuestras vidas y no sabemos cómo toman sus decisiones.

La Unión Europea está publicando leyes y recomendaciones para proteger a las personas y conseguir, entre otras cosas, que lo que hacen estos algoritmos no sea opaco (Council of the European Union, 2024). Se habla de auditarlos y se habla de transparencia, pero también es bien sabido que incluso en el caso de que pudiéramos obligar a las empresas a publicar el código fuente de sus algoritmos, una empresa puede saber cómo sacó al mercado (o a Internet) un algoritmo, pero a menudo no sabe cómo es en la actualidad, qué ha ido aprendiendo, qué prejuicios, sesgos y preferencias tiene en la actualidad. Son, además, cada vez más complejas estas cajas negras, y a menudo nos resulta ya imposible conocer los detalles de su funcionamiento, o cómo llegan a una determinada conclusión (Castelvecchi, 2016).

Por todo ello, las leyes europeas tratan de asegurar, entre otras cosas, que la decisión no queda exclusivamente en manos del algoritmo, sino que debe haber siempre una persona responsable que vigile las decisiones algorítmicas, y garantice que el proceso es correcto y que está libre de sesgos y errores (Council of the European Union, 2024; Ponce, 2024b).

Por tanto, no es solo la parte técnica la que debe preocuparnos. Tanto o más importante es conocer dónde pueden fallar las personas que usan el algoritmo, especialmente si son las responsables de garantizar su buen funcionamiento. Por muy bien que esté el algoritmo diseñado desde el punto de vista técnico, si induce errores humanos, o si está explotando algún sesgo humano o influyendo en las emociones o las decisiones humanas (ya sea de manera deliberada en su programación, o por error, o mal diseño), esto puede causarnos grandes problemas.

La investigación de la interacción humanos-máquinas es fundamental para nuestra supervivencia y nuestro bienestar en el mundo actual. Es además un campo que está en este momento en gran crecimiento, en donde se necesita mucha investigación psicológica. Hay en este momento muchos profesionales de otras áreas (legisladores, juristas, filósofos, sociólogos, ingenieros, matemáticos) trabajando en mejorar la ética de los algoritmos y en lograr que sean beneficiosos para la humanidad, pero los psicólogos debemos aportarles los datos que necesitan para poder realizar su trabajo.

Debemos investigar, por ejemplo, si determinada configuración por defecto propicia un mayor sesgo en las personas que interactúan con el algoritmo, por qué esta otra configuración debería llevar a una colaboración humanos-máquinas mucho más efectiva, por qué esta otra IA sería muy necesario probarla con determinadas modificaciones, antes de distribuirla en los hospitales, o en Internet.

Debemos investigar, también, si las personas encargadas de garantizar las buenas decisiones conjuntas con la IA de un hospital o un juzgado, serán capaces de hacerlo sin más, o si por el contrario, es necesario proporcionarles un asesoramiento especializado, y un protocolo de actuación que garantice su independencia y su motivación para poder redactar los informes pertinentes en caso de que detecten errores o sesgos en las decisiones y sugerencias del algoritmo. Todo esto es investigación psicológica y los psicólogos sabemos cómo hacerla. Sabemos que no es solo la máquina, sino la máquina en su interacción con la persona lo que hay que investigar.

Es un área en la que se está trabajando mucho en este momento, pero como podrán intuir, está aún casi todo por hacer. Cada pregunta que logramos contestar nos abre otras tantas preguntas, a cual más interesante, para futuras investigaciones.

Mostraré a continuación algunos experimentos en los que estamos trabajando actualmente en nuestro laboratorio, para ilustrar mejor lo que pretendo decir.

3.1. - Influencia de algoritmos de plataformas de ocio sobre decisiones importantes de las personas

Estamos actualmente muy acostumbrados a dejar que los algoritmos de las grandes tecnológicas decidan los libros y las noticias que leemos, la

música que escuchamos, las películas que vemos. Dejamos que tengan una gran relevancia en nuestras vidas. Delegamos la decisión y nos fiamos casi ciegamente de las recomendaciones del algoritmo de Amazon, el de Google, el de Spotify, o incluso, a la hora de hacer ciencia, de las recomendaciones del algoritmo de Research Gate, o el de Google, por mencionar solamente unos pocos. Creemos que nos conocen bien, que saben lo que nos interesa, que además velan por nuestro bienestar, y que, por tanto, nos conviene seguir sus consejos.

Pero dejando de lado la pregunta de hasta qué punto sus recomendaciones buscan realmente nuestro bienestar, o el beneficio de la empresa a la que pertenecen, dejando esa cuestión a los legisladores, que son los que pueden y deben velar por el buen uso de la IA, sí podemos como psicólogos hacernos preguntas de investigación, y hacer experimentos que son los que podrán proporcionar precisamente los datos que los legisladores necesitan para dictar leyes basadas en la evidencia. Podemos exigir una política y una legislación basadas en pruebas, pero para eso debemos ser capaces de proporcionar esas pruebas, averiguando qué aspectos de la IA benefician, y cuáles perjudican, al ser humano.

Algunos de esos datos que debemos proporcionar los psicólogos, son, por ejemplo: ¿pueden los algoritmos de IA influir en decisiones humanas importantes? Es decir, ya sabemos que influyen en decisiones de compra, en decisiones de qué música escuchar, qué película ver, qué noticias leer, incluso en si procede o no conceder una hipoteca a un cliente o una beca a un estudiante, pero en el supuesto de que esto no nos parezca suficientemente grave, ¿influirán también en nuestras decisiones más personales, como la de a quién votar en unas elecciones o incluso a quién proponer una cita romántica en una aplicación de citas vía internet?

Son preguntas empíricas y las pusimos a prueba en un trabajo reciente. La respuesta es: sí, las IAs pueden influir en decisiones humanas relevantes (Agudo y Matute, 2021). Utilizando, por motivos éticos, una IA ficticia que hacía recomendaciones a nuestros participantes sobre a quién votar en unas elecciones ficticias con políticos ficticios, los participantes acababan votando con más frecuencia a los candidatos que la IA les recomendaba como más compatibles con su modo de ser, de la misma manera que podía estar recomendándoles un coche o un libro.

También observamos en estos experimentos que en otro contexto muy diferente pero igualmente crítico, el de las decisiones de citas románticas vía internet, podíamos obtener resultados similares si explotábamos alguno de los sesgos cognitivos de los participantes. Una vez más, utilizando un contexto ficticio por cuestiones éticas, las personas podían pedir una cita *online* al candidato o candidata de su elección.

En este caso, los participantes fueron un poco más resistentes a la simple recomendación explicita por parte de la IA sobre a quién deberían solicitar la cita. Sin embargo, no tuvimos problema en explotar otros sesgos cognitivos para que solicitaran la cita al candidato (o candidata) diana de nuestro experimento. Sencillamente sustituimos en este experimento la recomendación explícita por otra mucho más sutil, que consistía simplemente en presentar más veces la fotografía del candidato o candidata diana que las de otros candidatos. Tal como indica el profesor Jaime Vila en su discurso de ingreso en la Academia:

"El rostro es posiblemente el estímulo que mayor información social y emocional proporciona… aporta muchos datos relevantes para la interacción social…, incluso las actitudes e intenciones amistosas u hostiles de las personas" (véase Vila, 2017, pág. 71).

Por ello, utilizamos fotografías de rostros de personas, y una de ellas la presentábamos con mayor frecuencia. De esta forma nuestro objetivo era simplemente explotar el sesgo de familiaridad, y efectivamente, logramos que los participantes votaran más al candidato cuyo rostro les presentamos con mayor frecuencia pues les resultaba más familiar que los candidatos controles.

Son ejemplos de investigaciones que estamos realizando en el momento actual, que nos animan a colaborar con otros profesionales de otras áreas de la psicología, y de otras disciplinas, y que creo que arrojan resultados preocupantes, y ponen de manifiesto que los psicólogos tenemos aún mucha investigación por hacer en estos temas y muchos datos que proporcionar aún a los legisladores.

+ + + + + + + + +

Pero obviamente no todo en la inteligencia artificial son aplicaciones dañinas diseñadas para influir en las decisiones humanas o para explotar nuestros

sesgos. Existen por supuesto infinidad de aplicaciones diseñadas para beneficio de la humanidad, que pueden realmente proporcionar grandes avances, pero que sin embargo pueden también en ocasiones contener errores y sesgos, como es el caso de muchos de los algoritmos de IA que se utilizan actualmente en sanidad. En esos casos, como veremos a continuación, necesitamos saber también cómo influyen esos posibles errores en las decisiones de las personas encargadas de trabajar con estas IAs.

3.2. - Influencia de los errores de los algoritmos en las decisiones humanas en sanidad

Un área en la que creo que todos podemos estar de acuerdo en que la IA está siendo utilizada para el bien de la humanidad, y que está mejorando, en general, la toma de decisiones humanas, es la sanidad. Los sistemas de ayuda a la decisión en el diagnóstico médico, por ejemplo, están demostrando ser de gran ayuda ya que permiten analizar muchos más casos y, en general, de manera bastante fiable.

Pero no son, ni mucho menos, 100% fiables, y es sabido que estos sistemas pueden contener sesgos y errores, que normalmente adquieren, o bien de sus programadores humanos, o bien durante su entrenamiento con bases de datos que también suelen contener sesgos y errores.

Por ello, la Unión Europea ha establecido que una persona tiene que ser siempre responsable de que no haya errores en la decisión final (Council of the European Union, 2024; Ponce, 2024b). Pero la pregunta que nos hacemos es: ¿Tiene esa persona realmente capacidad para vigilar el proceso, detectar los errores de la máquina e informar sobre ellos? Esto fue lo que nos preguntamos en una investigación que realizamos recientemente (Vicente y Matute, 2023). El objetivo era investigar cómo se comportan, y qué decisiones toman, las personas que están trabajando con estas IAs cuando tienen errores sistemáticos.

Para ello desarrollamos una tarea de diagnóstico médico utilizando una serie de muestras de tejido humano (que una vez más, son ficticias) que se presentaban a los voluntarios en la pantalla del ordenador. A la mitad les decíamos que cuando las muestras fueran más oscuras se trataba de un caso positivo de la enfermedad X y cuando fueran más claras se trataba de un caso

negativo; a la otra mitad le decíamos lo contrario. Las muestras se componían de una serie de pixeles más claros y más oscuros cuyas proporciones podíamos manipular fácilmente de manera que unas muestras resultaran muy claramente casos positivos o casos negativos de la enfermedad y otros casos eran más difíciles de clasificar.

En general, se trataba de una tarea sencilla que los participantes del grupo de control, que realizaban la tarea por sí mismos, sin ayuda de IA, realizaban correctamente en la mayoría de los casos.

Lo interesante es lo que ocurría con los participantes del grupo asistido por IA. La supuesta IA les mostraba, junto a cada muestra de tejido, su recomendación, sugiriendo clasificar la muestra como positiva o como negativa. La IA ofrecía casi siempre la recomendación correcta, pero tenía un sesgo. En un determinado tipo de muestras siempre se equivocaba, recomendando siempre la clasificación incorrecta: si era una muestra positiva, la IA recomendaba clasificarla como negativa, y viceversa.

El resultado fue que en esas muestras en las que la IA se equivocaba, los participantes del grupo experimental se equivocaban también, seguían la recomendación de la IA sin pensar por sí mismos, y cometían, por tanto, el mismo error. Estos errores no se daban en el grupo de control.

Es más, cuando posteriormente les decíamos que habíamos apagado la IA y que debían continuar realizando la tarea por sí mismos, siguieron cometiendo los mismos errores que había cometido la IA durante la fase anterior. Es decir, de alguna manera, la IA había actuado como un maestro en el que los participantes confiaron ciegamente, y aprendieron a reproducir tanto sus aciertos como sus errores, perpetuando estos incluso cuando la IA dejó de estar disponible.

Esto es grave. Y nos lleva además a un círculo del que nos puede costar salir: la IA aprende sesgos humanos, y posteriormente los transmite a la siguiente generación de humanos, que confía ciegamente y aprende de la IA, adquiriendo sus sesgos. Recordemos además que estas personas eran las que supuestamente debían garantizar que no hubiera errores en la decisión, pero si son entrenadas por la propia IA, pueden acabar simplemente heredando el mismo sesgo. Es más, incluso en los casos en que estas personas sean capaces de detectar el error, ¿podemos garantizar que tendrán la motivación y la confianza para informar del error, contradiciendo la recomendación de la IA?

Estos experimentos son ejemplos de cómo hay aún muchísima investigación que realizar, muchísimas preguntas por resolver, pues una vez que sabemos que las IAs heredan sesgos de las bases de datos humanas, y que posteriormente los trasmiten a la siguiente generación de humanos ante los que actúan como maestros incuestionables, se hace necesario investigar cómo reducir estos problemas, cómo conseguir que las personas encargadas de trabajar con las IAs sean más críticas y sean capaces no solo de detectar esos errores, sino también de informar sobre ellos, algo que también plantea numerosos retos interesantes.

En resumen, no pretendo realizar tampoco en este campo de la inteligencia artificial una revisión exhaustiva de todos los experimentos que se están realizando actualmente, pero creo que los ejemplos mencionados pueden servir para mostrar que podemos encontrarnos tanto con errores y sesgos involuntarios en la IA, como con otros que pueden estar programados de forma explícita para explotar sesgos y errores humanos, y conseguir que las personas se comporten de determinada manera, por ejemplo en elecciones democráticas, algo que debería preocuparnos bastante (Bond et al., 2012; Epstein y Robertson, 2015). Y estos sesgos no solo influyen en nuestras decisiones y comportamientos, sino que incluso aprendemos de ellos y podemos llegar a perpetuarlos.

En todos estos casos somos los psicólogos los que mejor conocemos el funcionamiento de la mente y la conducta humana y los que mejor podemos predecir situaciones y problemas que quizá los técnicos no pueden prever, o que quizá las empresas sí pueden prever, pero no siempre tienen interés en solucionar.

Por ello, es desde la psicología desde donde debemos hacer los experimentos y entregar los resultados a la sociedad, para que, entre todos, podamos lograr una IA que sea beneficiosa para la humanidad.

4. - Conclusiones

Ya nos avisó el gran escritor de ciencia ficción, Arthur C. Clarke (1962): "*Cualquier tecnología suficientemente avanzada es indistinguible de la magia*". La inteligencia artificial, desde luego, lo es, para la mayoría de nosotros. La tecnología actual es pura magia. Y lo que nos ocurre con la ciencia no es muy diferente de lo que ocurre en la tecnología en el momento actual, prolifera

también el pensamiento mágico, y mucha gente acude a las pseudociencias en busca de remedios y explicaciones, aunque no tengan ninguna base científica ni evidencia que las sustente. Se trata, por tanto, de dos grandes retos psicológicos de nuestro tiempo a los que debemos dar respuesta.

He intentado a lo largo del discurso precisamente mostrar que la pseudociencia es un problema psicológico y que los psicólogos deberíamos ser los primeros interesados en erradicarla de nuestra propia ciencia, y también de otras. Conocemos las limitaciones de la mente humana, y por tanto conocemos la utilidad, y la necesidad, del método científico para minimizar el impacto de los sesgos y evitar en lo posible que nos engañen productos y técnicas basados en ilusiones cognitivas.

También nuestra relación con la IA es un reto psicológico. Es verdad que la IA involucra a prácticamente todas las profesiones. Pero no olvidemos que para ser beneficiosa para la humanidad, necesita de investigación psicológica que aporte los datos y la evidencia que necesitan los legisladores para legislar a favor de las personas, y los ingenieros para realizar modelos que tengan en cuenta las limitaciones y las fortalezas de la mente humana, de modo que sea posible una interacción humanos-máquinas que potencie lo mejor de la humanidad y limite los riesgos.

Muchos profesionales necesitan los datos y los resultados de las investigaciones que los psicólogos debemos proporcionar (Ponce, 2024a). Aun siendo los políticos y los legisladores, los matemáticos y los ingenieros, los juristas y los jueces, quienes deberán solucionar muchos de los problemas que se están detectando tanto en relación con la IA como con las pseudociencias, no podrán hacerlo si no disponen de la evidencia científica, y proporcionarla es una tarea que nos corresponde a nosotros, pues en principio conocemos mejor que otras profesiones cómo pueden, tanto las pseudociencias como la IA, influir en las creencias, las decisiones, y la conducta humana.

Por ello, y para ir terminando, quisiera resaltar otra cita. Es una frase de Ramón López de Mántaras (2024), pionero de la investigación en inteligencia artificial en España. Decía recientemente, en relación a la IA, que "Invertir en educación (en la escuela y universidades) es nuestra mejor oportunidad para crear una sociedad que aproveche los beneficios de las tecnologías inteligentes mientras minimiza sus riesgos."

Estoy totalmente de acuerdo. Lo dijo en relación a la IA, y creo que exactamente lo mismo podríamos decir en relación a la ciencia, para crear

una sociedad que aproveche los beneficios de la ciencia mientras minimiza el uso de productos y técnicas pseudocientíficos.

En resumen, investigación y educación psicológica creo que son las mejores herramientas con las que contamos para hacer frente al pensamiento mágico tanto en ciencia como en tecnología, para conseguir una legislación basada en la evidencia, que beneficie a la humanidad, y una sociedad informada que sepa maximizar los beneficios y minimizar los riesgos.

Referencias

Agudo, U., & Matute, H. (2021). The influence of algorithms on political and dating decisions. PLOS ONE 16(4): e0249454. https://doi.org/10.1371/journal.pone.0249454

Allan, L. G., Siegel, S., & Tangen, J. M. (2005). A signal detection analysis of contingency data. Learning & Behavior, 33(2), 250–263. https://doi.org/10.3758/BF03196067

Blanco, F., Matute, H., & Vadillo, M. A. (2013). Interactive effects of the probability of the cue and the probability of the outcome on the overestimation of null contingency. Learning & Behavior, 41(4), 333-341. https://doi.org/10.3758/s13420-013-0108-8

Bond, R.M., Fariss, C.J., Jones, J.J., Kramer, A.D.I., Marlow, C., Settle, J.E., et al. (2012). A 61-million-person experiment in social influence and political mobilization. Nature, 489: 295–298. https://doi.org/10.1038/nature11421

Castelvecchi, D. (2016). The black box of AI. Nature, 538, 20-23.

Chow, J. Y., Colagiuri, B., & Livesey, E. J. (2019). Bridging the divide between causal illusions in the laboratory and the real world: the effects of outcome density with a variable continuous outcome. Cognitive Research: Principles and Implications, 4(1), 1-15. https://doi.org/10.1186/s41235-018-0149-9

Clarke, A. C. (1962). Profiles of the Future: An Inquiry into the Limits of the Possible. New York: Harper & Row.

Council of the European Union (2024). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf

Double, K. S., Chow, J. Y., Livesey, E. J., & Hopfenbeck, T. N. (2020). Causal illusions in the classroom: how the distribution of student outcomes can promote false instructional beliefs. Cognitive Research: Principles and Implications, 5(34), 1–20. https://doi.org/10. 1186/s41235- 020-00237-2

Echeburúa, E. (2023). Retos de los tratamientos psicológicos en la práctica clínica ante las nuevas demandas terapeúticas. En Academia de Psicología de España (Ed.): Psicología para un Mundo Sostenible III. Madrid: Pirámide.

Epstein R., & Robertson, R.E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. Proceedings of the National Academy of Sciences. https://doi.org/10.1073/pnas.1419828112

FECYT (2020). El uso y la confianza en las terapias sin evidencia científica. Ministerio de Ciencia e Innovación. Gobierno de España. https://www.fecyt.es/es/publicacion/el-uso-y-la-confianza-en-las-terapias-sin-evidencia-cientifica

FECYT (2022). Encuesta de percepción social de la ciencia y la tecnología en España (EPSCT). Ministerio de Ciencia e Innovación. Gobierno de España. https://doi.org/10.58121/msx6-zd63

Ferreira, J. L. (2015). Economía y pseudociencia. Díaz & Pons.

Ferrero, M., Garaizar, P., & Vadillo, M.A. (2016). Neuromyths in Education: Prevalence among Spanish Teachers and an Exploration of Cross-Cultural Variation. Frontiers in Human Neuroscience, 10:496. https://doi.org/10.3389/fnhum.2016.00496

Ferrero, M., Hardwicke, T. E., Konstantinidis, E., & Vadillo, M. A. (2020). The effectiveness of refutation texts to correct misconceptions among educators. Journal of Experimental Psychology: Applied, 26(3), 411–421. https://doi.org/10.1037/xap0000258

Freckelton, I. (2012). Death by homeopathy: issues for civil, criminal and coronial law and for health service policy. Journal of Law and Medicine, 19(3), 454-478.

García-Vera, M.P. (2023). La psicología en tiempos de pandemia: ¿Estamos siendo relevantes? En Academia de Psicología de España (Ed.): Psicología para un Mundo Sostenible III. Madrid: Pirámide.

Hannah, S.D. & Beneteau, J. L. (2009). Just tell me what to do: bringing back experimenter control in active contingency task with the command-performance procedure and finding cue density effects along the way. Canadian Journal of Experimental Psychology, 63(1), 59-73. https://doi.org/10.1037/a001340359

Hume, D. (1978). Treatise of human nature (edited by L.A. Selby-Bigge). London: Oxford University Press. (Originally published 1741).

Kahneman, D. (2011). Thinking fast and slow. New York: Farrar, Strauss and Giroux.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., … Nosek, B. A. (2014). Investigating variation in replicability. A "Many Labs" replication project. Social Psychology, 45(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Lilienfeld, S. O., Lynn, S. J., Namy, L., & Woolf, N. (2011). Psicología: Una introducción. Pearson.

Lobera, J., & Rogero-García, J. (2021). Scientific appearance and homeopathy. Determinants of trust in complementary and alternative medicine. Health Communication, 36(10), 1278-85. https://doi.org/10.1080/10410236.2020.1750764

López de Mántaras, R. (2024). Conocimientos de sentido común: el obstáculo de la IA en el camino hacia la inteligencia artificial general. The Conversation. https://theconversation.com/conocimientos-de-sentido-comun-el-obstaculo-de-la-ia-en-el-camino-hacia-la-inteligencia-artificial-general-235260

Martínez, N., Matute, H., Blanco, F., Barberia, I. (2024). A large-scale study and six-month follow-up of an intervention to reduce causal illusions in high school students. Royal Society Open Science, 11: 240846. https://doi.org/10.1098/rsos.240846

Matute, H., Blanco, F., & Díaz-Lago, M. (2019). Learning mechanisms underlying accurate and biased contingency judgments. Journal of Experimental Psychology: Animal Learning and Cognition, 45(4), 373–389. https://doi.org/10.1037/xan0000222

Matute, H., Blanco, F., Yarritu, I., Díaz-Lago, M., Vadillo, M. A., & Barberia, I. (2015). Illusions of causality: how they bias our everyday thinking and how they could be reduced. Frontiers in Psychology, 6, 888. https://doi.org/10.3389/fpsyg.2015.00888

Matute, H., Yarritu, I., & Vadillo, M. A. (2011). Illusions of causality at the heart of pseudoscience. British Journal of Psychology, 102(3), 392–405. https://doi.org/10.1348/000712610X532210

Moreno-Fernández, M.M., Blanco, F., & Matute, H. (2021). The tendency to stop collecting information is linked to illusions of causality. Scientific Reports, 11, 3942. https://doi.org/10.1038/s41598-021-82075-w

Mulet, J.M. (2015). Comer sin miedo. Booket.

O'Neil, C. (2016). Weapons of math destruction. Crown Publishing.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. Science, 349(6251)

Organización Médica Colegial de España (OMC) (s/f). Observatorio OMC contra las pseudociencias, pseudoterapias, intrusismo y sectas sanitarias. https://www.cgcom.es/observatorios/oppiss

Perales, J. C., Catena, A., Gonzalez, J. A., & Shanks, D. R. (2005). Dissociation between judgments and outcome-expectancy measures in covariation learning: A signal detection theory approach. Journal of Experimental Psychology: Learning Memory and Cognition, 31(5), 1105–1120. https://doi.org/10.1037/0278-7393.31.5.1105

Piejka, A. & Okruszek, Ł. (2020). Do you believe what you have been told? morality and scientific literacy as predictors of pseudoscience susceptibility. Applied Cognitive Psychology, 34(5), 1072-1082. https://doi.org/10.1002/acp.3687

Ponce, J. (2024a). El Reglamento de la Unión Europea de Inteligencia artificial de 2024, el derecho a una buena administración digital y su control judicial en España. Barcelona: Marcial Pons.

Ponce, J. (2024b). Inteligencia artificial, decisiones administrativas discrecionales totalmente automatizadas y alcance del control judicial: ¿Indiferencia, insuficiencia o deferencia? Revista de Derecho Público: Teoría y Método, 9: 9-56. https://doi.org/10.37417/RPD/vol_9_2024_2151

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), Classical conditioning II: Current research and theory (pp. 64-99). New York: Appelton-Century-Crofts

Rodríguez-Ferreiro, J., & Barberia, I. (2019). Believers in pseudoscience present lower evidential criteria. Scientific Reports, 11, 24352. https://doi.org/10.1038/s41598-021-03816-5

Rumelhart, D. E., & McClelland, J. L., & the PDP Research Group (1986). Parallel distributed processing: Explorations in the microstructure of cognition. Volume I: foundations & volume II: Psychological and biological models. Cambridge, MA: MIT Press.

Segovia, G., & Sanz-Barbero, B. (2022). 'It works for me': Pseudotherapy use is associated with trust in their efficacy rather than belief in their scientific validity. International Journal of Public Health 67, 1604594. https://doi.org/10.3389/ijph.2022.1604594

Shanks (2007). Associationism and cognition: Human contingency learning at 25. Quarterly Journal of Experimental Psychology, 60, 291-309. doi:10.1080/17470210601000581

Shanks, D. R., Pearson, S. M., & Dickinson, A. (1989). Temporal contiguity and the judgement of causality by human subjects. Quarterly Journal of Experimental Psychology, 41B, 139–159.

Torres, M. N., Barberia, I., & Rodríguez-Ferreiro, J. (2020). Causal illusion as a cognitive basis of pseudoscientific beliefs. British Journal of Psychology, 111(4), 840–852. https://doi.org/10.1111/bjop.12441

Tudela, P. (2023). De la psicología experimental a la neurociencia cognitiva. En Academia de Psicología de España (Ed.): Psicología para un Mundo Sostenible III. Madrid: Pirámide.

Vicente, L., & Matute, H. (2023). Humans inherit artificial intelligence biases. Scientific Reports, 13, 15737. https://doi.org/10.1038/s41598-023-42384-8

Vila, J. (2017). Emoción y supervivencia: Entre el corazón y el cerebro. En Academia de Psicología de España (Ed.): Psicología para un Mundo Sostenible, Vol. I. Madrid: Pirámide.

Vinas, A., Blanco, F., & Matute, H. (2023). Scarcity affects cognitive biases: The case of the illusion of causality. Acta Psychologica, 239, 104007. https://doi.org/10.1016/j.actpsy.2023.104007

Wasserman, E.A., & Neunaber, D.J. (1986). College students' responding to and rating of contingency relations: The role of temporal contiguity. Journal of the Experimental Analysis of Behavior, 46, 15-35

Yarritu, I., Matute, H., & Vadillo, M. A. (2014). Illusion of control: The role of personal involvement. Experimental Psychology, 61, 38-47. https://doi.org/10.1027/1618-3169/a000225.

Contestación del

Excmo. Sr. D. Enrique Echeburúa Odriozola

Discurso de contestación a la Académica Helena Matute Greño

Enrique Echeburúa Odriozola

Excelentísimo Sr. Presidente de la Academia de Psicología de España, excelentísimos miembros de la Academia, acompañantes, amigas y amigos todos. Me complace en extremo hacer la *laudatio* de Helena Matute en su incorporación como miembro de número de esta Academia y contestar a su discurso de ingreso en representación de los tres académicos proponentes: Pío Tudela, Jaime Vila y yo mismo. Trataré de explicar brevemente los motivos de esta complacencia.

Permítanme que aluda, en primer lugar, a los méritos de la nueva académica, que muestran una fructífera trayectoria científica, antes de comentar algunas notas acerca del contenido de su discurso de ingreso.

Respecto a su trayectoria académica, la Dra. Matute presentó su tesis doctoral en Psicología en la Universidad de Deusto en 1989, de donde es actualmente catedrática de Psicología Básica desde 1999. En cuanto a investigación, ha sido fundadora y dirige actualmente el Laboratorio de Psicología Experimental y el equipo de investigación del mismo nombre, que ha sido clasificado por el Gobierno Vasco como Equipo de Excelencia Investigadora. Por toda su trayectoria investigadora cuenta por el momento con 5 sexenios de investigación.

Sus líneas de investigación han experimentado una evolución con el transcurso del tiempo. Inicialmente se interesó por los procedimientos informatizados para poder investigar el aprendizaje y la memoria de manera experimental en el laboratorio. A finales de los años 90 empezó a realizar también los experimentos por medio de Internet, creando uno de los primeros laboratorios virtuales para la investigación *online* del aprendizaje y de la cognición. Más recientemente, su investigación se ha centrado en el estudio del aprendizaje causal, los sesgos e ilusiones cognitivas (especialmente el sesgo de causa-efecto) y la toma de decisiones, así como en las aplicaciones de todo ello a problemas actuales, tales como las pseudociencias y la desinformación.

Especial relevancia en su investigación actual tiene la posible manipulación de la conducta humana y la toma decisiones por medio de la inteligencia artificial (IA). Los estudios de la Dra. Matute y de su equipo se han orientado especialmente estos últimos años a la relación de las personas con los algoritmos de inteligencia artificial, concretamente en cómo puede influir esta

tecnología en las actitudes y la toma de decisiones, así como en la conducta humana. En concreto, no solo la inteligencia artificial hereda sesgos humanos, sino que el efecto contrario también se produce: las personas que trabajan con inteligencia artificial pueden heredar los sesgos de esta, de modo que podríamos estar iniciando un bucle complejo (incluso problemático) del que podría resultarnos difícil escapar, como lo ha mostrado recientemente en una publicación de *Scientific Reports* (2023) junto con Lucía Vicente, una de sus estudiantes de doctorado. Esta línea de trabajo ha suscitado un gran interés más allá del ámbito universitario.

En cuanto a publicaciones, ha sabido conjugar la investigación científica con la difusión de sus hallazgos. Es autora de 6 libros, entre ellos *Nuestra mente nos engaña*, publicado en 2019, y que ha tenido una gran resonancia en los medios de comunicación. A nivel bibliométrico, ha dirigido 14 tesis doctorales, tiene más de 100 publicaciones JCR y un índice h de 28 en la *Web of Science*.

Su línea de trabajo ha tenido una notable proyección internacional a nivel docente e investigador. Investigadora visitante en las universidades de Minnesota (EE. UU.), Queensland y Sídney (Australia) y Gante (Bélgica), ha impartido cursos de posgrado en numerosas universidades nacionales (UPV/EHU, UPNA, UNED, UIMP) e internacionales, entre otras en Lovaina, Bruselas, Santiago de Chile, Ciudad de México, Nueva York y Florida.

Respecto a sus cargos de gestión, ha sido presidenta de la Sociedad Española de Psicología Experimental (SEPEX, 2008-2010), así como de la Sociedad Española de Psicología Comparada (SEPC, 1995 y 2008). Ha sido colaboradora de la Dirección General de Investigación en la gestión, evaluación y seguimiento de los proyectos del Plan Nacional de I+D en el área de Psicología (2011-2014) y ha sido miembro del Comité Científico Asesor de la Fundación Española para la Ciencia y la Tecnología (FECYT, 2015-2020).

También ha desempeñado el cargo de editora asociada del *Quarterly Journal of Experimental Psychology* (QJEP, 1999-2004), que es la revista científica de la Experimental Psychology Society.

A la Dra. Matute se le han otorgado diversos reconocimientos por su trayectoria científica. Así, ha recibido el premio Universidad de Deusto–Grupo Santander en 2006; el Premio Ikerbasque a la trayectoria investigadora en 2020; y el premio Sanitarias en la categoría de Psicología en 2021. Es también académica de número de *Jakiunde*, la Academia de las Ciencias, las Artes y las Letras del País Vasco, desde 2017.

Comprometida con la transferencia del conocimiento y la divulgación científica, colabora habitualmente con los principales medios de comunicación y ha impartido numerosas conferencias al público general a través de universidades, así como a través de asociaciones ciudadanas, empresas e instituciones culturales (por ejemplo, Fundación Telefónica en Madrid, Centro de Cultura Contemporánea en Barcelona, Palacio Euskalduna en Bilbao, Teatro Victoria Eugenia y Museo de San Telmo en San Sebastián, La Domus en A Coruña, Baluarte en Pamplona, Museo de la Evolución Humana en Burgos, Ciudad de la Cultura en Santiago de Compostela, etc.).

Su labor de divulgación científica ha sido galardonada con el Premio Prisma al mejor libro de divulgación científica en 2002 por el libro *Adaptarse a Internet: Mitos y realidades sobre los aspectos psicológicos de la red*, y con el Premio JotDown -Donostia International Physics Center- al mejor artículo de divulgación científica en 2015 por el artículo *El capellán de la Virgen y el perro de Pavlov.*

En resumen, se trata de una profesora e investigadora comprometida con la ciencia y con la divulgación científica de temas socialmente relevantes. En concreto, actualmente se ocupa de un tema (los sesgos cognitivos en las pseudociencias y en la inteligencia artificial) que es una cuestión de interés indudable en todos los ámbitos y que puede aportar, por tanto, líneas de investigación novedosas para esta Academia.

Es una mujer entusiasta, comprometida con lo que hace y dispuesta a aceptar las exigencias de la plaza, con intención de implicarse activamente en el quehacer de la Academia y de colaborar con los miembros de esta institución. Como valor añadido, se incorpora por primera vez a la Academia una académica de la Universidad de Deusto, lo que supone incluir a esta institución de prestigio en las universidades presentes hasta la fecha en esta Casa.

Respecto al contenido del discurso de ingreso de la nueva académica, tiene como hilo conductor el problema de los sesgos cognitivos en las pseudociencias y en la inteligencia artificial (IA).

El análisis de las pseudociencias es muy sugerente. Por citar un ejemplo, en modo alguno exclusivo, en el ámbito de la psicología clínica y de la salud se practican muchas terapias que no están fundamentadas empíricamente y que, sin embargo, son de uso común. Al margen de que la inexistencia actual de pruebas sobre la eficacia de un determinado tratamiento no

debe ser tomada como prueba definitiva de su ineficacia, se considera en este entorno pseudociencia a cualquier intervención con una pretendida finalidad sanitaria que no tenga una evidencia científica (o que no la haya mostrado aún) que avale su eficacia para hacer frente con éxito a un trastorno mental o a un problema de salud. Entre las pseudoterapias, y sin ánimo de ser exhaustivo, están la hipnosis regresiva, la programación neurolingüística, el análisis transaccional, la magnetoterapia, la dieta macrobiótica, la sanación espiritual activa, el reiki, la terapia regresiva, las flores de Bach, las constelaciones familiares, el *focusing* o la terapia de crecimiento personal.

Si en el ámbito médico no puede prescribirse un fármaco sin una suficiente verificación de su eficacia y de la escasez de efectos adversos, no se entiende que en psicoterapia haya una especie de *barra libre* en que quienquiera puede hacer lo que guste, aun sin una formación profesional suficiente y sin una fundamentación empírica que lo justifique, como si las intervenciones psicosociales no pudieran tener efectos negativos. No es aceptable una especie de *curanderismo emocional* basado en pseudoterapias que no están fundamentadas en ensayos clínicos ni en revisiones sistemáticas o meta-analíticas. Y en particular la Sanidad Pública, sufragada con los impuestos de todos los contribuyentes, solo debería ofertar terapias efectivas y, en igualdad de condiciones, breves para acortar el sufrimiento del paciente y hacerlas accesibles en un tiempo razonable a los ciudadanos que las requieren.

Como señala la nueva académica, los sesgos cognitivos respecto a la eficacia de una terapia, en concreto la ilusión de causalidad, pueden darse, entre otras circunstancias, cuando hay una contigüidad temporal en una mejoría tras un tratamiento. Pero a veces dicha mejoría puede deberse a otros factores inespecíficos, como la aparición de una pareja, el cambio de actividad laboral, una mayor relación social o actividad lúdica, el paso curativo del tiempo o la aceptación de algo como inevitable (por ejemplo, en el caso de la muerte de un ser querido).

Por lo que se refiere a la IA, la inteligencia humana tiene una base biológica (se asienta en el cerebro humano, que se entreteje con millones de neuronas interconectadas), cuenta con una dimensión cognitiva de autoconciencia y emocional, desempeña una función adaptativa y muestra una capacidad única para la creatividad y la intuición. Por el contrario, la inteligencia artificial, al margen de su enorme utilidad como instrumento al servicio de los usuarios, es una creación humana basada en datos y algoritmos predefinidos, carece de conciencia y emociones y opera dentro de los límites de su programación

y de los datos disponibles. La IA, aun siendo extraordinariamente útil, carece de la flexibilidad cognitiva necesaria para adaptarse a las diferentes situaciones planteadas a un ser humano.

Se ha dicho que los algoritmos no tienen sesgos, pero los tienen porque los modelos de aprendizaje automático se elaboran con datos históricos que reflejan los prejuicios de la sociedad y pueden llevar a resultados discriminatorios en la toma de decisiones (por ejemplo, en el ámbito de la justicia penal, del diagnóstico médico o de la selección de personal).

A un nivel ético se requiere un control estricto. Los algoritmos de la IA pueden ser utilizados para manipular el comportamiento humano, como en la personalización de contenido en las redes sociales o en los mensajes publicitarios. Asimismo, la IA puede generar imágenes o vídeos realistas de personas sin su consentimiento, como en el caso de los *deepfakes*. Esto puede ser utilizado, por ejemplo, para crear contenido sexual explícito de personas que nunca dieron su permiso, violando su autonomía e intimidad.

Estas son algunas reflexiones que me ha suscitado a bote pronto el discurso de ingreso de la Dra. Matute. Es hora de terminar. No puedo ni debo concluir sin felicitar, en mi nombre, en el de los proponentes de su candidatura y en el de todos los miembros de esta Academia, a la nueva académica, Helena Matute Greño, por su trayectoria científica. Quiero darle nuestra bienvenida a esta Casa, de la que ya forma parte, invitarle a que la considere como propia y expresar nuestro deseo de una convivencia personal y profesional fructífera.

www.academiapsicologia.es