

Discurso de aceptación

19 de junio de 2025

Anthony G. Greenwald, galardonado en la categoría de Ciencias Sociales (XVII edición)

La cognición social implícita es un campo dentro de la psicología social que Mahzarin Banaji y yo introdujimos en un [artículo publicado en 1995](#) en *Psychological Review*. Concluimos aquel artículo afirmando nuestra expectativa de que este campo de investigación se desarrollaría rápidamente cuando dispusiéramos de un método para medir las diferencias individuales en los sesgos implícitos.

Unos años antes, Mahzarin y yo habíamos solicitado conjuntamente ayudas de la Fundación Nacional de Ciencias para investigar las actitudes implícitas y los estereotipos. En la Universidad de Washington, yo estaba poniendo a prueba cuatro técnicas que pensaba que podrían dar fruto. Una de ellas lo dio: el procedimiento que hoy se conoce como test de asociación implícita (o IAT por sus siglas en inglés: *Implicit Association Test*). El trabajo que condujo a ese primer IAT lo describí en un [capítulo autobiográfico de 2022](#).

Mientras desarrollaba los primeros test de asociación implícita, y [antes de la primera publicación sobre el IAT en una revista en 1998](#), creé versiones del procedimiento que eran aplicables en grupo y demostraron ser muy útiles en reuniones profesionales. Cuando quedó claro que estas demostraciones en grupo valían, Mahzarin también empezó a utilizarlas en sus presentaciones. Y, algo que fue aún más importante, enseguida empezó a utilizar el método IAT también en Yale, donde en su grupo de laboratorio estaba Brian Nosek, quien demostró ser una joya colaborando con nosotros dos en muchas investigaciones sobre la IAT. Antes de llegar a ocupar una cátedra en la Universidad de Virginia y establecer allí su propio laboratorio activo, Brian hizo el trabajo preliminar de abrir el sitio web educativo *Proyecto Implícito*, que en 2005 se convirtió en una organización sin ánimo de lucro 501(c)(3). Durante estos últimos veinte años, *Proyecto Implícito* ha sido el brazo didáctico del IAT, proporcionando gratuitamente gran variedad de mediciones autoaplicadas del IAT en numerosos idiomas.

Desde la creación del IAT, la investigación llevada a cabo con él y con métodos afines ha dado lugar a que los conocimientos sobre el sesgo implícito reciban gran atención y se utilicen en campos que van mucho más allá de la psicología, como los negocios, el derecho, la justicia penal, la medicina, la educación y las ciencias políticas. Este creciente corpus de investigación ha producido aplicaciones sociales útiles, pero lo que todavía no se ha conseguido es elaborar métodos eficaces para reducir el efecto discriminatorio de los sesgos implícitos; un efecto generalizado. Hoy en día, abordar este reto es muy prioritario.

Cuatro conclusiones demostradas están moldeando el trabajo que ahora se realiza en este campo para alcanzar dicho objetivo. Estas conclusiones han sido revisadas recientemente en publicaciones de [2020](#) y [2022](#).

Conclusión n.º 1: la mayoría de los sesgos implícitos, incluidos los relacionados con la raza, la edad y el sexo, están muy extendidos. Por ejemplo, los resultados de las mediciones del IAT que revelan una preferencia automática por los jóvenes en relación con las personas mayores están presentes en aproximadamente el 80 por ciento de los participantes en los estudios, entre ellos muchas personas que son mayores ellas mismas. Yo poseo ese sesgo implícito, además de otros tantos, como la preferencia automática por todo lo relativo a la «raza blanca» frente a la «raza negra», preferencia sobradamente observada, ya que la encontramos en aproximadamente el 60 por ciento de la mayoría de las muestras grandes.

Conclusión n.º 2: los resultados de las mediciones de sesgo implícito del IAT arrojan una clara correlación con actitudes y comportamientos discriminatorios; y aunque estas correlaciones no sean de gran magnitud, las actitudes y conductas discriminatorias que predicen son socialmente significativas debido al gran número de personas que las poseen y en las que operan sin que sean conscientes de ello.

Conclusión n.º 3: a pesar de la gran inversión hecha en «formación en materia de prejuicios implícitos» para empleados de compañías, hospitales, departamentos de policía de grandes ciudades y muchas otras empresas e instituciones, no hay pruebas contrastables de que estos esfuerzos formativos hayan reducido los prejuicios implícitos de las personas a quienes se imparten.

Conclusión n.º 4, tal vez la más interesante y al mismo tiempo la que plantea más dificultades: hace unos 10 años, los investigadores de la inteligencia artificial empezaron a descubrir que los resultados de sus intentos de utilizar la IA para automatizar las decisiones de contratación o promoción muchas veces son erróneos. Aunque se trate de seleccionar a los candidatos más dignos de ser contratados, las selecciones de los programas de IA generativa suelen ser discriminatorias por razón de sexo, raza o edad.

Cuando me enteré de que los [defensores de la IA estaban encontrando sesgos discriminatorios](#) en su uso de la IA generativa para la selección de personal, supuse que los investigadores en IA encontrarían rápidamente formas de reparar los sesgos de sus modelos y que podrían adaptar sus soluciones para ayudar a quienes toman las decisiones: seres humanos con sesgos implícitos.

Dado que, por desgracia, esto aún no ha sucedido, ahora estoy buscando otras posibilidades para reducir el sesgo implícito en la toma de decisiones.

Esta versión escrita de mi discurso, incluye un [suplemento](#) donde comento mi deuda con varias decenas de profesores, colegas y estudiantes que contribuyeron sustancialmente al desarrollo de mi carrera y de la investigación que el Premio Fronteras del Conocimiento de este año reconoce.

Estoy muy agradecido a la Fundación BBVA, a su jurado y a sus nominadores por incluir mi trabajo entre los revolucionarios avances reconocidos con el premio de Ciencias Sociales de este año.