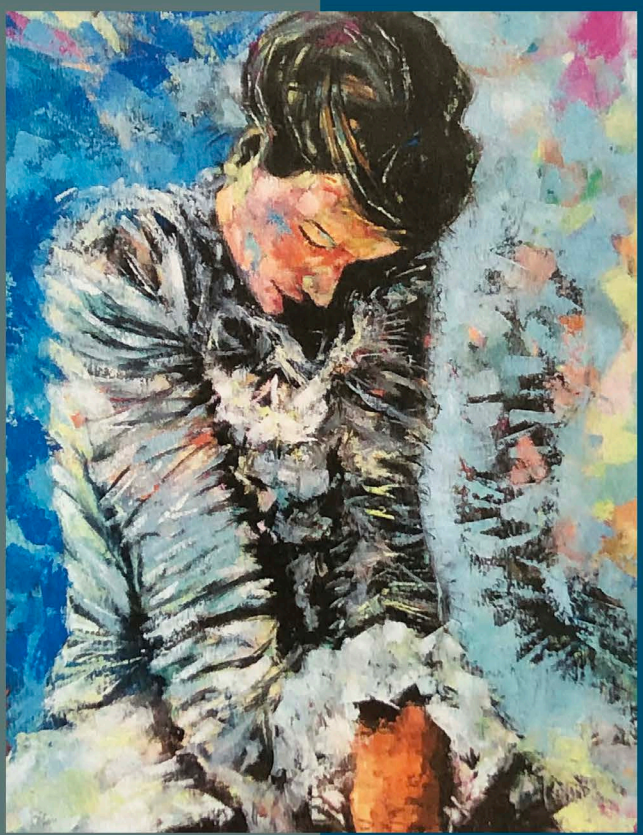
Razonamiento jurídico y ciencias cognitivas



Federico José Arena / Pau Luque / Diego Moreno Cruz

EDITORES

SERIE INTERMEDIA DE TEORÍA JURÍDICA Y FILOSOFÍA DEL DERECHO N30

Centro de Investigación en Filosofía y Derecho

Universidad Externado de Colombia

135 Años

FEDERICO JOSÉ ARENA
PAU LUQUE
DIEGO MORENO CRUZ
(Editores)

RAZONAMIENTO JURÍDICO Y CIENCIAS COGNITIVAS

UNIVERSIDAD EXTERNADO DE COLOMBIA

Razonamiento jurídico y ciencias cognitivas / María Laura Manrique [y otros] ; Federico José Arena, Pau Luque, Diego Moreno Cruz (editores). -- Bogotá : Universidad Externado de Colombia. 2021.
298 páginas ; 24 cm. (Intermedia de Teoría Jurídica y Filosofía del Derecho ; 30)

Incluye referencias bibliográficas.

ISBN: 9789587906417

1. Argumentación jurídica 2. Filosofía del derecho 3. Interpretación del derecho 4. Teoría del derecho I. Arena, Federico José, editor II. Luque, Pau, editor III. Moreno Cruz, Diego José, editor IV. Universidad Externado de Colombia V. Título VI. Serie

340.1 SCDD 15

Catalogación en la fuente -- Universidad Externado de Colombia. Biblioteca. EAP.
julio de 2021

ISBN 978-958-790-641-7

Calle 12 n.º 1-17 Este, Bogotá
Teléfono (57 1) 342 0288
publicaciones@uexternado.edu.co
www.uexternado.edu.co

Primera edición: julio de 2021

Pintura de cubierta: *Calamitas*, Magdalena Ana Rosso, óleo sobre lienzo, 50 x 40 cms., 2016
Diseño de cubierta: Departamento de Publicaciones
Corrección de estilo: Santiago Perea Latorre
Composición: Precolombi EU-David Reyes
Impresión y encuadernación: DGP Editores S.A.S.
Tiraje de 1 a 1.000 ejemplares

Impreso en Colombia
Printed in Colombia

V.
Acerca de la relevancia de las investigaciones sobre sesgos implícitos para el control de la decisión judicial

FEDERICO JOSÉ ARENA*

La categorización social (*i.e.*, la clasificación de personas en grupos sobre la base de determinadas características [TAJFEL, 1978]) es un componente básico de nuestro modo de pensar y actuar (LAKOFF, 1987). Asimismo, es una herramienta indispensable en el funcionamiento del derecho, puesto que las normas generales asocian una solución normativa a una clase de casos identificada, en parte, mediante referencia a una categoría de personas (ALCHOURRÓN y BULYGIN, 1971). No obstante, el uso de categorías sociales no está exento de dificultades. En ámbito jurídico se exige a los jueces que contrasten los efectos perjudiciales derivados de clasificaciones basadas en características sensibles (o sospechosas), referidas a grupos que han sufrido discriminación en el pasado (SABA, 2009). Numerosos ordenamientos jurídicos, con mayor o menor éxito, han introducido dispositivos para reducir la discriminación producida por categorizaciones sospechosas[1].

Sin embargo, desde hace unos años ha surgido una preocupación por un tipo de discriminación diferente, cuyos efectos no podrían ser combatidos mediante los instrumentos conocidos, puesto que estos últimos resultarían efectivos únicamente contra formas explícitas de discriminación. Esta preocupación se refiere a formas de discriminación producidas por creencias y/o actitudes que, de manera inadvertida para sus portadores, inciden sobre el modo en que clasifican a los demás y reaccionan frente a ellos. En efecto, un cúmulo considerable de trabajos en esos campos han apoyado la tesis según la cual es bastante común que las personas

---

* Doctor europeo en Filosofía del Derecho y Bioética Jurídica de la Universidad de Génova (Italia). Ha sido investigador posdoctoral en la Universidad Bocconi (Milán, Italia). Actualmente es investigador adjunto del Consejo Nacional de Investigaciones Científicas y Técnicas–CONICET (Argentina), profesor adjunto de Filosofía y Lógica Jurídica de la Universidad Blas Pascal (Argentina) y director de la revista *Discusiones*. Agradezco a Flavia Carbonell y a Pau Luque por haber comentado y criticado una versión anterior de este trabajo.

1 Por ejemplo, en Argentina pueden verse, entre otras, INADI (2005) y las leyes nacionales n.º 23.592 sobre medidas para evitar impedimentos arbitrarios al ejercicio de los derechos y garantías constitucionales, n.º 26485 de protección frente a la violencia contra las mujeres y n.º 26618 de matrimonio igualitario.

atribuyan, sin advertirlo, determinados rasgos a los miembros de ciertos grupos y asuman, también sin advertirlo, determinadas actitudes frente a esos grupos y sus miembros (Greenwald y Hamilton Krieger, 2006). Las creencias y actitudes de este tipo son frecuentemente denominadas "sesgos implícitos". Si bien más abajo me referiré a algunas imprecisiones de la expresión, los sesgos implícitos parecen representar un desafío para la lucha contra la discriminación puesto que sugieren la posibilidad de que las personas actúen de manera sesgada y traten a los demás de manera discriminatoria aun sin darse completamente cuenta de ello[2]. Es en virtud de estas posibles consecuencias que algunos teóricos del derecho han promovido la formulación de una agenda pública para lidiar con la incidencia de los sesgos implícitos en distintas áreas institucionales, incluyendo la decisión judicial (Jolls y Sunstein, 2006).

En efecto, investigaciones llevadas a cabo en ámbito jurídico pretenden mostrar que también los funcionarios encargados de aplicar la ley están sometidos a la influencia de sesgos implícitos. La categorización social cumple funciones tanto en la interpretación de los textos normativos como en la identificación de los hechos relevantes del caso. En esos dos ámbitos, los jueces deben llevar a cabo razonamientos que involucran la percepción de sujetos y su inclusión en grupos. Las categorías usadas pueden ser conscientes y, por lo tanto, estar bajo el control del juez; pero también, se afirma, pueden ser reflejo de sesgos implícitos. Así, la categorización puede influir tanto en la justificación externa de la premisa normativa como en la justificación externa de la premisa fáctica. En el primer caso, porque la elección de una determinada interpretación puede haber sido provocada por las propias categorías del juez. Es decir, entre dos interpretaciones posibles, el juez elegirá la que sea consistente o coherente con ellas (Arena, 2016). En el segundo caso, las categorías pueden determinar la manera en

---

2 En palabras de Jennifer Saul: "Todo esto debería perturbarnos profundamente. Pues significa que nuestros juicios están siendo dramáticamente injustos, aun cuando no lo sean de manera intencional. Tratamos a los miembros de grupos estigmatizados de manera incorrecta, incluso si deseamos con desesperación tratarlos correctamente. Aún más, que todo esto suceda contribuye a que ese trato injusto se perpetúe" (Saul, 2013: 246; trad. propia).

que el juez concibe la explicación de ciertos hechos y, por lo tanto, influir tanto en el grado de plausibilidad otorgado a ciertos elementos probatorios como en el grado de verificación conferido a ciertos enunciados fácticos (Coloma, 2010).

Por ejemplo, respecto de la interpretación, algunas investigaciones han buscado mostrar que los sesgos pueden incidir en el modo en que los jueces asignan significado a las disposiciones normativas. En un caso, se solicitó a un grupo de jueces que evaluaran el significado de una hipotética ley, destinada a despenalizar la tenencia de marihuana cuando existiera una declaración jurada de un médico donde conste que el acusado posee la marihuana por necesidad médica. El problema interpretativo que se proponía era si esa ley cubría también el caso de alguien que, no teniendo la declaración jurada al momento del arresto, la conseguía con posterioridad. De acuerdo con ese estudio, los rasgos del hipotético acusado inciden implícitamente en la interpretación. Así, los jueces interpretaron restrictivamente la ley cuando el acusado tenía alrededor de 19 años y usaba la marihuana para combatir convulsiones, y la interpretaron de manera extensiva cuando tenía 55 años y la usaba para mitigar los dolores de un cáncer de huesos (Wistrich *et al.*, 2015).

Los ejemplos son también numerosos respecto del razonamiento probatorio. El problema suele ser más agudo cuando se trata de delitos contra la integridad sexual. Tan es así que el Estatuto de Roma de la Corte Penal Internacional establece que "[l]a credibilidad, la honorabilidad o la disponibilidad sexual de la víctima o de un testigo no podrán inferirse de la naturaleza sexual del comportamiento anterior o posterior de la víctima o de un testigo", y que "no se admitirá pruebas del comportamiento sexual anterior o ulterior de la víctima o de un testigo"[3].

En todos estos casos, se afirma, la posibilidad de que sesgos implícitos influyan en la decisión del juez aumenta el efecto perjudicial de la discriminación. Por un lado, porque es posible que los jueces decidan de manera discriminatoria aun sin darse completamente cuenta de que lo están haciendo. Por otro lado, porque el impacto discriminador que puede

---

3 Véase Asensio *et al.* (2010: 83–112) para numerosos ejemplos de jurisprudencia argentina en la materia, donde se reflejan tales sesgos en el razonamiento probatorio.

tener el proceso de categorización social es mayor cuando se trata de sesgos implícitos intrainstitucionales. En efecto, si bien los sesgos implícitos de los operadores jurídicos suelen reproducir formas de categorización ya existentes en la sociedad a la que pertenecen, el hecho de que tales sesgos determinen ciertos resultados institucionales puede profundizar (o incluso zanjar) la diferenciación social discriminatoria.

Estas investigaciones parecen apoyar entonces la afirmación de que las decisiones judiciales, como gran parte de otras acciones que involucran la categorización social, pueden sufrir la incidencia de sesgos implícitos. Sin embargo, desde mi punto de vista, queda todavía por precisar exactamente qué rendimiento pueden tener estas investigaciones a los fines de elaborar estrategias de control de la decisión judicial. Es decir, creo que es necesario aún avanzar en esclarecer de qué manera las investigaciones sobre sesgos implícitos mejoran o agregan elementos a los mecanismos ya existentes de control de la decisión judicial. Por ello, en este trabajo quisiera detenerme en el modo en que las investigaciones sobre sesgos permitirían identificar decisiones judiciales equivocadas, concentrándome sobre todo en las decisiones interpretativas, *i.e.*, en las decisiones judiciales acerca del significado de las disposiciones normativas[4]. Esta indagación exige, desde mi punto de vista, abordar preliminarmente algunas dificultades conceptuales.

Una de esas dificultades es cierta ambigüedad en la noción de sesgo implícito, entre sesgo como causa y sesgo como resultado. Para elucidar esta ambigüedad, propondré en el siguiente apartado (1) una mínima cartografía conceptual y terminológica, para asegurar un entendimiento compartido de los términos que serán usados. Ello es necesario ya que, por un lado, hay más formas problemáticas de categorización social que los sesgos implícitos, como los estereotipos y los prejuicios, y, por otro lado, hay más formas de deficiencia cognitiva que la involucrada en la categorización social.

A continuación de ese recorrido conceptual, abordaré dos discusiones suscitadas entre quienes se dedican al tratamiento de los sesgos implícitos tanto a nivel empírico como filosófico. La primera, a la que dedicaré el

---

4 Si bien haré también algunas referencias a las decisiones en ámbito probatorio.

apartado 2, es acerca de la validez científica de los experimentos que han sido llevados adelante en ciencias cognitivas para mostrar la existencia de sesgos implícitos y sobre cuya base se propone la agenda institucional mencionada más arriba. La segunda, analizada en el apartado 3, es acerca de las conclusiones escépticas sobre cuestiones de epistemología normativa que, se afirma, surgen de las investigaciones sobre sesgos implícitos.

Reconstruiré brevemente los extremos de las dos discusiones mencionadas. Mi intención no es resolver cada una de ellas, sino identificar algunas conclusiones que pueden extraerse a partir del modo en que tales controversias son planteadas. En este sentido, la conclusión más conservadora que propondré es que, incluso si los sesgos implícitos son un componente de nuestro modo de razonar y comportarnos, es necesario distinguir entre, por un lado, el hecho de que la acción sea causada por una actitud o estado mental inconsciente o fuera del control del agente y, por otro lado, el hecho de que el resultado de esa acción sea correcto o incorrecto según cierto criterio normativo. La distinción es importante, no solo por razones terminológicas (es frecuente encontrar en la literatura un uso impreciso del término "sesgo" para hacer referencia, sin distinguir, tanto a la causa como al resultado de la acción) sino, más que nada, porque permite, por una parte, advertir los compromisos conceptuales de las afirmaciones acerca de sesgos implícitos y, por otra parte, precisar la incidencia o relevancia de la investigación sobre sesgos implícitos respecto de la decisión judicial. A la presentación de estas tesis estará dedicado el apartado 4.

Finalmente, en el apartado de conclusiones y consideraciones finales introduciré el problema relativo al tipo de responsabilidad, si es que cabe, que es posible atribuir a quien actúa movido por un sesgo implícito.

1. LA MADEJA CONCEPTUAL

Por lo general, mediante un estereotipo se asocia a una categoría de personas, por el solo hecho de pertenecer a esa categoría, o bien un determinado rasgo o características (estereotipos descriptivos) o bien determinado rol (estereotipos normativos). Es decir, se atribuye un rasgo B o rol C a todos los miembros de una categoría en virtud de que, en cuanto poseen la propiedad A, pertenecen a esa categoría (Oakes *et al.*, 1994). El rasgo o el rol atribuidos pueden ser positivos o negativos (*i.e.*, los cumbieros son

violentos, los asiáticos son buenos en matemáticas, las madres deben ser amas de casa, los hombres deben proveer alimentos). Los estereotipos que atribuyen un rasgo suelen ser analizados como generalizaciones. Ello en cuanto, en el mejor de los casos, se trata de afirmaciones acerca de la existencia de una correlación estadística entre poseer la propiedad adscrita a los miembros de un grupo y el hecho de pertenecer a ese grupo (SCHAUER, 2003 y APPIAH, 2000).

Un prejuicio, en cambio, implica una valoración o actitud respecto de los miembros de cierto grupo que puede estar o no asociada a un estereotipo. Es decir, la actitud positiva o negativa frente a los miembros de un grupo puede depender de considerar que tales personas, en cuanto pertenecen al grupo, poseen una cierta propiedad o estar basada directamente en la pertenencia al grupo.

De este modo, es posible distinguir, por un lado, una reacción cognitiva frente a miembros de un grupo (cuando se estereotipa) y una reacción actitudinal (a partir de prejuicios). Ahora bien, se trata de reacciones típicamente explícitas, en dos sentidos. Por un lado, porque el portador del estereotipo o del prejuicio es consciente de ello y, por otro lado, porque ambos tipos de reacción se manifiestan en el comportamiento. Ahora bien, lo que mostrarían las investigaciones recientes es que ambos tipos de reacciones pueden también ser implícitas, donde implícitas se asocia en la mayoría de los casos a la falta de consciencia por parte del portador y, en consecuencia, a la falta de control de sus efectos. Es decir, es posible que las personas reaccionen de manera estereotipada o prejuiciosa sin advertir que dicha reacción ha sido, precisamente, impulsada por un estereotipo o prejuicio. Un punto importante es entonces precisar el sentido de "implícito" para evitar la ambigüedad apenas señalada relativa a "explícito". El punto que caracteriza a los sesgos implícitos es la inconsciencia y su manifestación en acciones no intencionales, es decir, en acciones no destinadas a manifestar ese estado mental o actitud. Así, un sesgo implícito, tal como una atribución inadvertida de un rasgo o rol a una persona por el mero hecho de pertenecer a un grupo (sesgo asociado a estereotipo) o tal como cierta actitud frente a una persona por el mero hecho de pertenecer a un grupo (sesgo asociado a prejuicio), puede dar lugar a acciones sesgadas. Por ejemplo, la tendencia a favorecer el propio grupo de pertenencia es quizás uno de los sesgos más extendidos. En un test aplicado a jueces de

Minnesota, se les pidió que determinaran la indemnización que correspondía otorgar en un caso hipotético de daño producido por el derrame de líquidos químicos peligrosos en un lago. Cuando el hipotético demandado era de Minnesota, la indemnización era la mitad de la indemnización impuesta cuando el demandado era de un Estado diferente (WISTRICH y RACHLINSKI, 2018: 99).

Ahora bien, los estudios que han investigado la incidencia, en nuestras acciones, de procedimientos mentales inadvertidos, han identificado una amplia variedad de sesgos que no se refieren únicamente a categorías sociales. Se trata, como decía más arriba, de otros tipos de sesgos cognitivos. En este sentido, los sesgos en general constituyen estrategias o recursos que el sujeto aplica de manera inconsciente cuando procesa información. La intervención de procedimientos cognitivos automáticos o inadvertidos es inevitable, pues es imposible que todas las tareas cognitivas de la mente sean conscientes. Sin embargo, no es infrecuente que tales procesos resulten sesgados. Entre los posibles sesgos suelen indicarse, entre otros, los siguientes:

(a) Sesgo de disponibilidad: los sujetos suelen considerar como más probable el acaecimiento de un evento cuando les resulta familiar, es decir, cuando pueden recordar sucesos similares.

(b) Sesgo del anclaje: los sujetos suelen estimar una magnitud a partir de un valor inicial, por lo que, a diferente valor inicial, diferente estimación. Algo parecido sucede con los precios de los autos. El comprador sabe que el precio de lista sirve como punto de inicio para la negociación de una rebaja, pero sabe que el pedido de una reducción tiene un límite. El anclaje también parece afectar el razonamiento de los jueces. En un caso se solicitó a jueces indicar el monto que correspondía otorgar como indemnización en virtud de daños producidos por trato inapropiado en el lugar de trabajo. Algunos jueces recibieron, entre el material probatorio, un testimonio de la víctima donde contaba que había visto en televisión que a una mujer le habían concedido una indemnización de 415.300 dólares por un hecho similar. Los jueces que recibieron ese testimonio concedieron una indemnización promedio de 50.000 dólares, mientras que la indemnización promedio concedida por quienes no lo recibieron fue, en cambio, de 6.000 dólares (WISTRICH y RACHLINSKI, 2018: 93).

(c) Sesgo de la confirmación: los sujetos suelen filtrar información (o evidencia) de manera tal que solo impacta en su conocimiento aquella que confirma su posición inicial[5].

Aquí no me interesa el detalle de estos otros tipos de sesgos, pero hay un punto respecto de ellos que resulta relevante para lo que diré más adelante. En cada uno de estos ejemplos es necesario distinguir dos aspectos que suelen ser confundidos bajo el rótulo de sesgo. Por un lado, el hecho de que la acción haya sido causada por un procedimiento automático (sesgo como causa) y, por otro lado, el hecho de que el resultado de la acción (ya sea una decisión, una nueva creencia, etc.) resulte equivocado (sesgo como resultado). Así, una creencia acerca de una magnitud producida por un sesgo de anclaje bien puede, no obstante su origen, ser correcta (*i.e.*, no sesgada). Lo que, en caso contrario, vuelve equivocada a la creencia resultante del sesgo como causa no es el hecho de haber sido provocada por el sesgo, sino el hecho de que transgrede un criterio, explícito, que permite identificarla como equivocada. Es decir, la identificación misma de una creencia sesgada implica la existencia de un criterio de corrección explícito, independiente del sesgo que la causa. Dicho en otras palabras, es posible que la acción haya sido motivada por un sesgo como causa y, sin embargo, que el resultado no sea sesgado, *i.e.*, si satisface los criterios de corrección. Incluso más, la existencia misma de esos criterios de corrección es indispensable para la identificación de los sesgos como causas. Solo si contamos con criterios para identificar resultados sesgados es posible identificar los sesgos-causa, puesto que estos no se expresan de otro modo (eso es precisamente lo que los vuelve implícitos). En la próxima sección regresaré sobre este punto.

2. La epistemología de lo implícito

La preocupación por la incidencia de factores inconscientes en los comportamientos no es ciertamente nueva. Sin embargo, la reciente ola de experimentos en ciencias cognitivas ha conferido cierta legitimidad a estas

---

5 Para una lista más nutrida, véase De la Rosa Rodríguez y Sandoval Navarro, 2016 y Muñoz Aranguren, 2011.

investigaciones. Legitimidad que la epistemología empirista suele negar al psicoanálisis. Además, estas investigaciones recientes tienen también la particularidad de presentar un desafío a estructuras conceptuales tradicionales, como, por ejemplo, la teoría analítica de la acción humana que tradicionalmente ha incluido a la intención como elemento definicional del concepto de acción (Von Wright, 1963). Ello en cuanto, a partir de estas investigaciones se afirma que "los agentes no siempre tienen un control consciente e intencional de los procesos de percepción social, de formación de impresiones y de evaluación que motivan sus acciones" (Greenwald y Hamilton Krieger, 2006: 946; trad. propia).

Las investigaciones que apoyan la tesis de la existencia de sesgos implícitos son numerosas y los experimentos llevados a cabo acumulan ya una gran cantidad de iteraciones y réplicas (ibíd.). Quizás los dos ejemplos siguientes pueden ser ilustrativos de este tipo de investigación. Por un lado, el experimento que llevaron a cabo los psicólogos Peters y Ceci acerca de los sesgos en los procedimientos para la evaluación y publicación de artículos académicos. Por otro lado, el test conocido como sesgo del tirador. En el primer caso, Peters y Ceci enviaron a las más importantes revistas de psicología artículos que ya habían publicado, pero con nombres falsos y sin afiliación a una institución prestigiosa. Solo unas pocas revistas detectaron el plagio, y la gran mayoría (casi el 90%) los rechazaron, señalando serias deficiencias metodológicas (Peters y Ceci, 1982). En el segundo caso, se trata de estudios donde se solicita a los sujetos que disparen (virtualmente) si y solo si ven en las imágenes proyectadas a un sujeto armado. La mayoría de los sujetos tendieron a percibir un arma cuando quien sostenía un objeto ambiguo era una persona negra y un teléfono cuando era una persona blanca (Correll *et al.*, 2007).

Más en general, los experimentos más conocidos y que han adquirido cierta estabilidad son el Test de Asociación Implícita (iat, por sus siglas en inglés) y los estudios denominados del detonador/disparador emocional (*affective-priming*).

El iat es un método destinado a medir la velocidad con la que los sujetos reaccionan frente a estímulos que exigen asociar un par de categorías de personas con adjetivos positivos y negativos. En una versión de este test se exige a los participantes que establezcan, usando el teclado de la computadora, si la imagen de un rostro proyectada en el centro de la pantalla

es un rostro de una persona negra o blanca, y si una palabra proyectada en el centro de la pantalla es una palabra positiva o negativa. Las opciones (blanco/negro-positivo/negativo) están ubicadas, por pares opuestos, en los ángulos superiores izquierdo y derecho de la pantalla. La tecla "E" es usada para ubicar la imagen o palabra del lado izquierdo y la letra "I" para ubicarla del lado derecho. Por lo general se comienza por asociar rostros, se continúa asociando palabras, luego rostros y palabras a blanco/positivo y negro/negativo y luego a blanco/negativo y negro/positivo. El sistema mide el tiempo de cada respuesta.

Así, lo que propone el test es que las diferencias en la velocidad de asociación están vinculadas a un sesgo, en el sentido de que, si implícitamente el sujeto asocia una categoría con un valor positivo, ello se reflejará en que su reacción será más rápida cuando tenga que asociar esa categoría con un valor positivo y será más lenta cuando tenga que asociar una categoría diferente con ese mismo valor. En breve, a mayor velocidad en la reacción, mayor incidencia de un sesgo. Por ejemplo, asociar más rápidamente los rostros y las palabras cuando el par es blanco/positivo y negro/negativo implica poseer un sesgo desfavorable hacia los negros.

El test de *affective-priming* también se basa en la medición de la velocidad de la reacción, pero se diferencia del IAT en que busca activar el sesgo mediante la proyección de una imagen que, si bien impacta en sus sentidos, no es advertida por el sujeto que participa del experimento. Así, si luego de la proyección de esa imagen subliminal el sujeto asocia más rápidamente la imagen subsiguiente (esta sí, percibida conscientemente) con un adjetivo positivo, quiere decir que posee un sesgo positivo respecto de la imagen proyectada de manera subliminal. Y viceversa en el caso de que la asociación se produzca más rápidamente con un adjetivo negativo. Por ejemplo, en algunos experimentos se proyectaban de manera subliminal imágenes de personas blancas y de personas negras e inmediatamente después se solicitaba que se clasificaran ciertos objetos.

Ambos tipos de test pueden ser encontrados y, algunos, realizados en línea, y constan de un altísimo número de réplicas y participantes en línea (véase: https://implicit.harvard.edu/implicit/argentina/). Sin embargo, la validez de tales investigaciones ha dado lugar a una encendida controversia. Aquí me referiré brevemente a la discusión entre, por un lado, Gregory Mitchell y Philip Telock (MITCHELL y TETLOCK, 2006) y, por otro, Samuel

Bagenstos (BAGENSTOS, 2007). Si bien esta controversia gira sobre varios ejes, entre ellos el uso retórico de "ciencia" e "investigación científica", a continuación me concentraré en la crítica que, a partir de un conjunto de estándares sobre la validez de investigaciones científicas, Mitchell y Tetlock dirigen contra los estudios sobre sesgos implícitos, y la respuesta de Bagenstos a esa crítica. Mediante la reconstrucción de la controversia no pretendo, ni estoy capacitado para, zanjar la discusión. Me interesa, en cambio, intentar identificar algunas conclusiones acerca de cómo identificar sesgos implícitos que, si bien son independientes de la resolución de la controversia, son relevantes para el problema que aquí nos interesa.

Mitchell y Tetlock lamentan que los autores que insisten en la necesidad de una agenda pública contra la incidencia de los sesgos implícitos no hayan considerado antes los problemas relativos a las credenciales científicas de las investigaciones en las que apoyan esa agenda[6]. En efecto, según estos autores, existen elementos suficientes para dudar de tales credenciales. Para probar esta afirmación, Mitchell y Tetlock proponen evaluar las investigaciones sobre sesgos implícitos a partir de cuatro estándares de validez científica: (1) Validez de constructo, (2) Validez interna, (3) Validez de la conclusión estadística y (4) Validez externa (MITCHELL y TETLOCK, 2006: 1056).

El primer estándar, de validez del constructo, exige que el objeto inobservable de una investigación haya sido adecuadamente operacionalizado o traducido en una variable que pueda ser manipulada y medida a los fines de testear las hipótesis. El problema con las investigaciones sobre sesgos implícitos es, según los autores, que las variables propuestas como medida de los sesgos son ambiguas. Por ejemplo, el IAT propone como medida de los sesgos la variación en la velocidad de asociación. Sin embargo, esa variable puede ser interpretada como consecuencia de otros procesos alternativos. Además, debería existir cierta convergencia en los resultados de mediciones alternativas del mismo constructo, y ello no ha sido el caso entre IAT y *affective-priming*.

El segundo estándar, de validez interna, exige eliminar explicaciones alternativas que puedan dar cuenta del mismo fenómeno. Es decir, los

---

6 Véase, para una defensa de esta agenda, JOLLS y SUNSTEIN, 2006.

experimentos deberían producirse en contextos donde estén controladas variables alternativas e igualmente explicativas de la acción discriminatoria (Mitchell y Tetlock, 2006: 1032-1033). Los autores señalan una lista de interpretaciones y/o explicaciones alternativas tales como: diferencias en la familiaridad con los distintos grupos sociales, amenaza de estereotipo (*i.e.*, el fenómeno, identificado inicialmente por Claude Steele, que se produce cuando las personas, sabiéndose miembros de un grupo estereotipado, tienden a "confirmar la profecía" encapsulada en el estereotipo [Steele, 2010]), compasión o culpa producidas por la situación de alguno de los grupos involucrados en el experimento, conocimiento del prejuicio de otros o existente en la sociedad, entre otros.

El tercer estándar (de validez de la conclusión estadística), por su parte, requiere que las mediciones estadísticas posean cierta precisión. Sin embargo, mientras que se ha demostrado que muchos factores diferentes inciden en el tiempo de reacción (como la flexibilidad cognitiva, asimetrías respecto de la familiaridad del estímulo, entre otras), los desarrolladores del IAT asumen que en todos los casos indican sesgos frente a los estímulos.

Finalmente, el estándar de validez externa exige que la extensión, a contextos no controlados, de las conclusiones obtenidas en el contexto artificial de laboratorio se apoye en investigaciones aplicadas. Dicho en otras palabras, no está justificado asumir sin más que los resultados de laboratorio permiten predecir comportamientos en el mundo real (Mitchell y Tetlock, 2006: 1034). Por ejemplo, si bien el IAT fue también administrado a jueces en diferentes ocasiones y el resultado promedio fue similar al del resto de las personas que participaron, no siempre el resultado del IAT predijo correctamente la reacción de jueces frente a casos hipotéticos que ejemplificaban los sesgos (Wistrich y Rachlinski, 2018: 101).

Frente a estas críticas, destinadas a cuestionar la validez científica de las investigaciones sobre sesgos implícitos, Samuel Bagenstos defiende la agenda pública dirigida a contrarrestar los efectos de los sesgos implícitos. Su argumento principal es que la crítica, supuestamente epistemológica, de Mitchell y Tetlock refleja en realidad un desacuerdo normativo con los defensores de esa agenda pública. Más precisamente, según Bagenstos,

> … muchas de las críticas de Mitchell y Tetlock a la investigación sobre sesgos implícitos descansan, no tanto en una base científica, sino en un conjunto de

> presuposiciones normativas acerca de qué clase de discriminación debería ser prevenida y castigada por el ordenamiento jurídico. En particular, las críticas se apoyan en una concepción demasiado estrecha, basada en nociones de culpa individual, según la cual el derecho debería prohibir únicamente la discriminación que sea el resultado de un *animus* consciente e irracional […] Sin embargo, quienes promueven el uso del derecho para combatir los sesgos implícitos no asumen esa concepción. Por el contrario, conciben a la discriminación como un problema social que –refleje o no la "culpa" de un individuo discriminador– tiene efectos sistemáticamente dañinos sobre las oportunidades de vida de los miembros de ciertos grupos sociales (Bagenstos, 2007: 479-480; trad. propia).

Por ejemplo, como vimos, Mitchell y Tetlock afirman que procesos diferentes a la hostilidad racial pueden producir modificaciones en la velocidad de reacción medida por el IAT. Por su parte, Bagenstos sostiene que esta crítica es en realidad normativa, pues de acuerdo con nuestro autor, y a diferencia de lo que parecen sostener Mitchel y Tetlock, no existe ninguna diferencia relevante, desde una perspectiva anti-discriminación, entre que el resultado del sesgo refleje hostilidad o cualquiera de las otras causas señaladas por estos autores. Lo que importa es que el comportamiento y las actitudes limitan las oportunidades de los miembros de grupos minoritarios. Así, la falta de familiaridad con miembros de grupos minoritarios en ciertos contextos laborales hará que quienes pertenecen a grupos mayoritarios mantengan sus reacciones aversivas, y ello producirá efectos negativos, reflejen o no hostilidad. De este modo, siempre según Bagenstos, tampoco es determinante la supuesta falta de satisfacción del cuarto estándar, pues, por ejemplo, si fuese cierto que las personas en contextos de laboratorio actúan de manera sesgada impulsadas por la amenaza de estereotipo, es probable que esa amenaza también incida fuera del laboratorio, y, en ese caso, lo que cuenta, de nuevo, es que producirán efectos sesgados, afectando los derechos de los miembros de grupos minoritarios. Es esta afectación lo que importa a una política de protección antidiscriminatoria (Bagenstos, 2007: 485).

Así, para Bagenstos, la crítica que está detrás del trabajo de Mitchell y Tetlock es en realidad una crítica normativa, a saber, que los defensores de la agenda pública contra los sesgos implícitos no han todavía ofrecido una sólida justificación de la respuesta a actos de discriminación sistémica

y no individuales. El punto sobre el que presiona Bagenstos es ciertamente relevante y, como señalé en la introducción, exige elaborar una teoría acerca de la relación entre sesgos y responsabilidad individual. Pero sobre este último punto solo diré algunas palabras en el último apartado. Lo que aquí me interesa resaltar es que de la misma defensa que Bagenstos hace de la agenda sobre sesgos implícitos resulta posible extraer como conclusión la necesidad de distinguir entre el sesgo como causa y el sesgo como resultado. No solo para evitar caer en una discusión meramente terminológica[7], sino también para identificar qué es lo que criticamos de una acción o creencia cuando afirmamos que es sesgada. Bagenstos es incluso más enfático acerca de que lo verdaderamente relevante no es que las acciones hayan sido producidas por un especial estado mental implícito, sino el hecho de que el resultado de estas sea discriminatorio. Para que esta exigencia normativa de Bagenstos tenga sentido es necesario que poseamos criterios explícitos e independientes que permitan evaluar la corrección o incorrección del resultado de la acción. En el caso que nos interesa, que permita establecer si posee o no efectos discriminatorios.

3. LA DUDA ESCÉPTICA

El segundo problema teórico que quisiera abordar aquí se refiere a cuáles son las consecuencias filosóficas, en ámbito epistemológico, que han de ser extraídas a partir de los resultados de las investigaciones sobre sesgos implícitos. Más específicamente, quisiera analizar el argumento según el cual la demostración de la existencia de sesgos, que no advertimos ni controlamos, pero que influyen en nuestro modo de percibir y reaccionar frente a ciertas categorías de personas, implica inevitablemente una conclusión escéptica general acerca de la validez de nuestro conocimiento.

Me concentraré aquí en la formulación que de tal argumento propone Jennifer Saul (Saul, 2013). Saul sostiene que la evidencia empírica acerca de la existencia de sesgos en el modo en que percibimos a los demás pone en jaque una buena porción de la empresa científica. Gran parte de las afirmaciones científicas se apoyan en el testimonio, es decir, la gran mayoría

7 Sospecha que surge leyendo el trabajo de Bagenstos.

de los enunciados que consideramos verdaderos y justificados se apoyan en el trabajo y las afirmaciones de otras personas. Cuando conocemos a través del testimonio, el conocimiento es adquirido a partir de otros y la adquisición de ese conocimiento exige inevitablemente categorizar a la persona sobre cuyo testimonio nos apoyamos. El problema, señala el argumento escéptico, es que esta última actividad, *i.e.*, la inclusión de otros en categorías, está afectada por la incidencia de sesgos implícitos. Si ello es así, entonces las investigaciones sobre cómo los sesgos afectan nuestro razonamiento ofrece fuertes razones para dudar de la validez del conocimiento que hemos adquirido.

Suelen distinguirse tres ámbitos filosóficos en los que puede incidir el escepticismo. En primer lugar, el escepticismo metafísico acerca de X es aquel que niega que existan hechos del tipo X. En segundo lugar, el escepticismo semántico acerca de un universo de discurso afirma que los enunciados de ese discurso carecen de valor de verdad (o, en algunas versiones, que son todos falsos). En tercer lugar, el escepticismo epistémico acerca de un universo de discurso es el que sostiene la tesis según la cual no es posible establecer el valor de verdad de los enunciados de ese discurso. Saul no está interesada en los dos primeros tipos de escepticismo (además, como señalaré más abajo, está constreñida a negarlos). En efecto, la autora no niega que existan hechos respecto de los cuales nuestros enunciados sobre el mundo empírico puedan ser contrastados. Además, Saul admite que algunos de nuestros enunciados sobre el mundo son verdaderos (*i.e.*, los que afirman la existencia de sesgos). Su escepticismo es, tal como ella misma lo señala, eminentemente epistémico, su tesis es que no podemos estar seguros acerca del valor de verdad que atribuimos a un conjunto de enunciados. Veamos cuáles son estos enunciados y las razones que Saul señala a favor de su conclusión escéptica.

Las investigaciones sobre sesgos implícitos, sostiene Saul, muestran que tenemos razones para dudar de nuestras facultades destinadas a la búsqueda del conocimiento, y ello es así puesto que gran parte de nuestro conocimiento está basado en el testimonio. Los sesgos implícitos inciden de manera inconsciente en el modo en que percibimos y clasificamos a las personas. Dado que el conocimiento basado en el testimonio proviene de otras personas, los sesgos influyen también en el modo en que percibimos y clasificamos a las personas a partir de cuyo testimonio construimos nuestro

conocimiento. En consecuencia, existe un alto riesgo de que los enunciados que creemos apoyados en evidencia (testimonial) se encuentren afectados por estos sesgos y, por lo tanto, no podemos estar seguros (*i.e.*, tenemos buenas razones para dudar) de que efectivamente conozcamos su valor de verdad (Saul, 2013: 244). De acuerdo con Saul, gran parte de nuestro conocimiento lo hemos obtenido de esta manera. Las consecuencias de este argumento son incluso más devastadoras, pues llegados a este punto no podemos siquiera identificar qué porción de ese conocimiento se produjo o no de manera sesgada, puesto que no podemos ya recordar el origen exacto de cada enunciado, *i.e.*, no podemos recordar el testimonio exacto a partir del cual hemos establecido la verdad o falsedad de cada enunciado. Como conclusión deberíamos dudar de un gran número de los enunciados que actualmente consideramos verdaderos y justificados.

Saul sostiene que la forma de escepticismo epistémico que defiende es distinta de otras formas tradicionales o más conocidas del mismo tipo de escepticismo como, por ejemplo, la tesis según la cual no tenemos evidencia que permita descartar la hipótesis de que no somos más que cerebros en un balde. Las diferencias entre este tipo de escepticismo y el de Saul son de fuerza, de alcance y pragmáticas. Con relación a la fuerza, Saul señala que mientras el escepticismo tradicional solo nos confiere razones para *dudar* de la verdad de todos nuestros enunciados sobre el mundo externo, el escepticismo producto de los sesgos implícitos nos ofrece buenas razones para *creer* que no podemos confiar en nuestras capacidades epistémicas. En segundo lugar, mientras el escepticismo del cerebro en el balde ofrece razones para dudar de que *todos* nuestros enunciados acerca del mundo exterior son falsos, el escepticismo de los sesgos implícitos alcanza solo un subconjunto de tales enunciados, *i.e.*, los que involucran el testimonio como fuente de conocimiento. Es decir, mientras el escepticismo tradicional es global, puesto que todos nuestros enunciados pueden tener un valor de verdad distinto del que les atribuimos, el escepticismo de Saul es no global (podemos estar seguros del valor de verdad de algunos enunciados). Sin embargo, dada la difusión que, según la misma Saul, el testimonio tiene en el modo en que adquirimos conocimiento, su escepticismo no sería local (*i.e.*, reducido a un conjunto definido y acotado de enunciados) sino casi global, es decir, tenemos razones para dudar del valor de verdad que hemos atribuido a gran parte de nuestros enunciados acerca del mundo.

Tercero, mientras que los efectos pragmáticos del escepticismo tradicional se agotan una vez que nos levantamos del sillón filosófico, el escepticismo de los sesgos implícitos nos exige ponernos en acción. Es decir, del hecho de que seamos cerebros en un balde no se sigue ninguna razón para cambiar el modo en que vivimos, pues el engaño está tan bien pertrechado que nuestras vidas podrían transcurrir igual en ese caso. En cambio, el escepticismo de los sesgos nos exige actuar para evitar que nuestro modo equivocado de clasificar a las personas incida en lo que creemos conocer y, así, incida en nuestras acciones (Saul, 2013: 244). Este último punto exige entonces que estemos atentos para evitar que los sesgos puedan incidir en nuestras evaluaciones de los demás; según Saul,

> … el problema empieza a hacerse evidente cuando nos preguntamos a nosotros mismos cuándo deberíamos preocuparnos por la influencia de los sesgos implícitos en nuestros juicios. La respuesta es que deberíamos preocuparnos cada vez que tengamos que considerar una afirmación, un argumento, una sugerencia, una pregunta, etc. de una persona cuyo grupo social de pertenencia estamos en condiciones de reconocer (ibíd.: 250; trad. propia).

Ahora bien, me parece que la argumentación que propone Saul tiene algunos problemas que suelen afectar al escepticismo. Problemas que provienen de lo que llamaré el efecto del grifo abierto. Este efecto suele incidir en los escepticismos, supuestamente no globales, sino casi globales. Ello es así porque este tipo de escepticismo, si bien afirma que gran parte de los enunciados de un ámbito del discurso están afectados por el aguijón escéptico, presupone necesariamente que al menos un enunciado de ese ámbito del discurso está libre del defecto, pues es precisamente sobre ese enunciado que se apoya el argumento escéptico. En este sentido, es conocido el análisis que Dworkin lleva a cabo del escepticismo interno, casi global, respecto de los enunciados morales (Dworkin, 2011: 23-97). Un ejemplo de este tipo de posición es la que afirma que todos nuestros enunciados morales son falsos puesto que carecemos de libre albedrío. Dworkin señala que este escepticismo, a primera vista global, es necesariamente casi global, pues depende de la verdad de al menos un enunciado moral, a saber, el enunciado contrafáctico según el cual si poseyéramos libre albedrío, tendríamos algunos deberes morales. El escepticismo de Saul está también constreñido

a ser casi global, puesto que su tesis, acerca de la incidencia de los sesgos implícitos en la adquisición de conocimiento basado en testimonio, presupone que somos capaces de adquirir conocimiento acerca de la existencia de sesgos implícitos. Sin embargo, y aquí está el efecto del grifo abierto, las razones por las que hemos de dudar de nuestro conocimiento basado en el testimonio se extienden también al conocimiento sobre la existencia de sesgos implícitos. A continuación intento mostrar por qué ello es así.

Una vez que aceptamos que tenemos fuertes razones para dudar de nuestro conocimiento basado en el testimonio, puesto que la adquisición de ese conocimiento exige la categorización de personas y las investigaciones en ciencias cognitivas muestran que sufrimos de sesgos cuando llevamos a cabo tal categorización, tenemos buenas razones para dudar también de las investigaciones en ciencias cognitivas, puesto que los experimentos de laboratorio llevados cabo por tales científicos incluyen la categorización de personas. Quizás Saul podría argüir que, a diferencia de otros defensores de la existencia de sesgos implícitos, ella no niega que seamos capaces de advertir la existencia de tales sesgos y de intentar evitarlos. Es decir, gracias a los resultados de las investigaciones sobre sesgos, ahora podemos (estamos obligados a) estar atentos al modo en que adquirimos conocimiento a través del testimonio y evitar así la influencia de sesgos. Dicho con otras palabras, el aguijón escéptico afecta a todo el conocimiento anterior a estas investigaciones, pero podemos ser optimistas (o, al menos, no tan pesimistas) acerca del futuro, pues ahora sabemos que hemos de estar alerta. El problema con esta respuesta es que los experimentos llevados a cabo en ciencias cognitivas se realizaron antes de obtener su resultado, es decir, antes de saber que nuestra categorización de las personas está precedida por sesgos. Esto nos pone frente al dilema siguiente: o bien aceptamos que hemos logrado conocimiento (*i.e.*, aceptamos que tenemos buenas razones para considerar como verdaderos ciertos enunciados) antes de saber que sufrimos de sesgos implícitos, pero entonces hemos de aceptar que hay otros enunciados, además de los que identifican sesgos, en los que podemos confiar; o bien defendemos un escepticismo casi global y por lo tanto tenemos buenas razones para dudar que hayamos logrado obtener conocimiento antes de saber que los sesgos implícitos influyen en nuestra categorización, pero entonces tenemos buenas razones incluso para dudar de los resultados de las investigaciones sobre sesgos implícitos.

Al igual que en la sección anterior, me parece que la conclusión que puede ser extraída a partir del recorrido apenas realizado y del dilema al que nos ha llevado el escepticismo de Saul es que la identificación de sesgos implícitos, mediante las investigaciones en ciencias cognitivas, presupone criterios de corrección de enunciados sobre el mundo que no pueden reducirse al criterio simple de que no hayan sido causados por sesgos. Criterios que, por lo tanto, bien podemos usar para evaluar el resultado de otros procedimientos cognoscitivos, *e.g.*, de otros enunciados considerados verdaderos y justificados. Es decir, si queremos evitar el segundo cuerno del dilema (el escepticismo global), debemos aceptar que contamos con criterios para identificar enunciados que ofrecen conocimiento y que, sobre esa base, podemos identificar como sesgado un resultado que se aparte de ese criterio. Dicho con otras palabras, la identificación de sesgos-resultado exige que contemos con criterios de corrección independientes a la existencia misma del sesgo implícito-causa.

4. VOLVIENDO AL ÁMBITO JURÍDICO

Veamos cómo el análisis llevado adelante hasta aquí puede ser de utilidad para abordar la incidencia que las investigaciones sobre sesgos implícitos pueden tener respecto del control de la actividad judicial, en particular, del conjunto de acciones a través de las cuales los jueces determinan la existencia y el contenido del derecho. Hay una limitación inicial a la incidencia que los sesgos poseen en la decisión judicial, y es que la actividad judicial se produce en un contexto de decisión extendido en el tiempo. Es decir, los jueces cuentan con la posibilidad de deliberar con mayor o menor detenimiento acerca de qué decisión tomar. En contextos decisorios de este tipo, la incidencia de los sesgos, si bien no desaparece, tiende a disminuir. En este sentido, los psicólogos suelen distinguir entre dos formas de tomar decisiones, a saber, uno intuitivo o automático y otro deliberativo o reflexivo. Esta diferencia suele, a su vez, expresarse en términos de la distinción entre el Sistema 1 y el Sistema 2, respectivamente. Así, el Sistema 1 sería aquel que involucra la intuición y las emociones, produciendo decisiones más veloces y apoyadas, por lo general, en asociaciones automáticas entre conceptos. El Sistema 2, en cambio, involucra capacidades intelectuales de orden superior y produce decisiones razonadas, apoyadas en la deducción

lógica o el razonamiento analógico. Es por ello que las decisiones a través del Sistema 2 insumen mayor tiempo (véase KAHNEMAN, 2012). Obviamente, como señalan varios autores, esta estricta distinción entre dos tipos de formas de tomar decisiones es, en cierto sentido, una ficción puesto que ambos sistemas suelen entrelazarse al momento de tomar una decisión concreta (WISTRICH y RACHLINSKI, 2018: 90).

Ahora bien, incluso en los casos en que los sesgos afecten decisiones tomadas en contextos donde predomina el Sistema 2 cabe preguntarse qué aporte puede esperarse de las investigaciones sobre sesgos para el control de la decisión judicial. A través del recorrido analítico propuesto en este texto he intentado mostrar la ambigüedad teórica de la noción de sesgo implícito. A veces, con la expresión "sesgo implícito" se hace referencia al estado mental que provoca, de manera inadvertida, una determinada acción, mientras que otras veces la expresión es usada para hacer referencia al resultado de ciertas acciones. Desde mi punto de vista, una vez sacada a la luz esta ambigüedad es posible advertir que la identificación misma de la existencia de un sesgo como causa presupone que contamos con criterios explícitos, e independientes de esa causa, para determinar si la acción (su resultado) es sesgado. Es decir, criterios destinados a identificar cuándo el resultado de la acción se aleja de cierto criterio y, en el caso de la categorización social, cuándo el resultado de la acción es discriminatorio.

En el ámbito de la decisión judicial, tales criterios son ofrecidos por las convenciones o métodos interpretativos y por los estándares de prueba. Esto es, la primera exigencia que se impone a un juez, al momento de identificar el significado de las disposiciones normativas, es que respete las convenciones interpretativas, *i.e.*, que interprete correctamente. Asimismo, se exige al juez que, al momento de evaluar los elementos probatorios, decida según el estándar de prueba correspondiente, ya sea el de más allá de toda duda razonable, el de prueba preponderante, mejor explicación u otro. De este modo, una interpretación o una valoración de la prueba aparecerán sesgadas precisamente cuando se aparten de esos criterios. Si ello es así, entonces también en ámbito judicial la identificación misma del sesgo presupone la existencia de un criterio normativo, en este caso métodos interpretativos y estándares de prueba. Pero si ya contamos con esos criterios para el control de la decisión judicial, entonces la inclusión de los sesgos implícitos no agregaría nada. Dicho con otras palabras, existiendo

criterios de corrección de la decisión judicial, lo que importa es que esos criterios sean respetados. Incluso si la decisión fue motivada por un sesgo como causa, si el resultado no es sesgado, *i.e.*, si satisface los criterios de corrección, entonces no existirá un sesgo como resultado.

Por supuesto, se puede todavía argüir que una decisión jurídicamente correcta puede ser, de todos modos, discriminatoria. Pero incluso en este caso, el criterio para determinar la naturaleza discriminatoria de la decisión, que ahora ya es distinto del criterio de corrección jurídico, será un criterio explícito, idéntico a los criterios para identificar acciones producto de las formas tradicionales de categorización social explícita y problemática, como estereotipos y prejuicios. Lo que estoy intentando señalar es que tanto el criterio de corrección jurídica como un posible criterio de discriminación extra-jurídico son criterios explícitos e independientes de la creencia o actitud que causó la decisión.

El argumento anterior podría ser tachado de formalismo ingenuo y cientifismo desmedido pues los criterios de corrección interpretativa y probatoria no están siempre disponibles y, por lo tanto, las consideraciones precedentes pueden parecer algo simplistas desde el punto de vista de la teoría de la interpretación o de la teoría de la prueba. Enfrentar esta observación exige formular una aclaración. Ciertamente existen ámbitos de indeterminación interpretativa y probatoria, más o menos amplios según el ordenamiento jurídico. El argumento anterior no niega esto, ni necesita hacerlo, simplemente señala que, allí donde esos criterios son determinados, los hallazgos sobre sesgos implícitos no agregan nada a las herramientas provenientes de los análisis de las formas explícitas de categorización social, tales como estereotipos y prejuicios. Ahora bien, ¿qué sucede en aquellos casos de indeterminación? La primera impresión es que la respuesta sería que es precisamente en aquellos casos donde no hay criterios de interpretación claros o estándares de prueba precisos donde la decisión se vuelve discrecional y puede verse afectada por un sesgo y, por lo tanto, producir resultados discriminatorios. En este punto vale la pena introducir una distinción propuesta por Ricardo Caracciolo entre diferentes sentidos en que una decisión puede ser discrecional:

(a) Discreción fuerte: es la situación en la que el ordenamiento jurídico no provee ninguna respuesta al caso que debe ser decidido y por lo tanto

no hay ninguna decisión jurídica correcta que respetar. Por ejemplo, en los casos de laguna normativa.

(b) Discreción reglada: es la situación en la que el ordenamiento jurídico otorga al juez la facultad para tomar una decisión, dentro de un conjunto limitado de decisiones alternativas. Por ejemplo, para la determinación de las penas dentro de un mínimo y un máximo.

(c) Discreción conferida: es la situación en la que el ordenamiento jurídico delega en el juez qué decidir caso por caso (Caracciolo, 2009)[8]. Por ejemplo, cuando el ordenamiento expresamente le indica al juez que decida de manera "razonable" o según su "leal saber y entender".

En estos tres casos, y más allá de que en el caso (b) el juez deba decidir dentro de un marco de opciones, no contamos con criterios interpretativos o probatorios para identificar una decisión correcta e incorrecta. Y, por lo tanto, si no hay decisión correcta desde el punto de vista de algún criterio, entonces ninguna decisión es sesgada, al menos desde el punto de vista de criterios jurídico-interpretativos. Obviamente, se puede todavía argüir que la identificación de la decisión sesgada, dado que no hay criterios jurídicos, depende ahora de criterios que permiten determinar su naturaleza discriminatoria. Pero, de nuevo, se tratará también de un criterio explícito, idéntico a los criterios para identificar acciones producto de las formas tradicionales de categorización social explícita y problemática, como estereotipos y prejuicios.

Todo esto parece llevarnos a un dilema, donde ambos cuernos incluyen la tesis de la irrelevancia de los sesgos implícitos. Es decir, o bien poseemos criterios jurídicos para identificar decisiones sesgadas, pero en ese caso nos alcanza con los criterios jurídicos y la investigación sobre sesgos es irrelevante; o bien no poseemos criterios jurídicos y por lo tanto la identificación de decisiones sesgadas exige recurrir a criterios extra-jurídicos, pero en ese caso se tratará de criterios explícitos que, de nuevo, vuelven irrelevante la cuestión sobre la incidencia de los sesgos implícitos.

En realidad creo que no es complicado escabullirse de este dilema. En efecto, quienes defienden una agenda pública, incluido el control de

---

8 Para otras distinciones entre diferentes formas de discreción, véase Dworkin, 1984 (1978) y Fletcher, 1984.

decisiones judiciales, a partir de las investigaciones sobre sesgos, no están preocupados por la decisión individual, sino por el efecto agregado de decisiones sesgadas. En este sentido, creo que el criterio que surge de esa agenda es la exigencia de igualdad transversal caso a caso. Por ejemplo, cuando se lamenta que un juez, "pocas semanas después de imponer una condena de seis meses por agresión sexual a un estudiante universitario blanco de Stanford, imponga una condena de tres años a un inmigrante mexicano por un delito similar" (WISTRICH y RACHLINSKI, 2018: 88; trad. propia).

En estos casos, siempre hay un criterio normativo externo que permite identificar el resultado sesgado, pero no se trata ya de un criterio relativo a la decisión individual, sino de la exigencia de igualdad entre decisiones. Este tipo de criterio permitiría identificar decisiones sesgadas incluso en aquellos casos en los que no hay criterios jurídicos para hacerlo. Si bien el criterio de igualdad ya alcanzaría para imponer esta exigencia y, por lo tanto, los sesgos implícitos podrían, de nuevo, aparecer como irrelevantes, el punto de estas investigaciones es que ofrecen una explicación y por lo tanto un diagnóstico y un remedio, para los casos en que los resultados agregados de las decisiones judiciales muestren un resultado desigual entre grupos sociales, en especial, cuando perjudican a grupos vulnerables o que han sido discriminados históricamente. Decisiones disparatadas en casos símiles permite sospechar de la incidencia de un sesgo.

CONSIDERACIONES FINALES

Generalmente se considera que la actividad (*i.e.*, las decisiones) de los jueces resulta controlada por mecanismos institucionales específicamente previstos a tal efecto, en particular los distintos recursos procedimentales. Estos recursos permiten que un órgano judicial, superior al que tomó la decisión, controle la corrección de la decisión del tribunal inferior. Se trata, en consecuencia, de un tipo de control vertical. Los límites de este control se manifiestan cuando se trata de órganos judiciales supremos, es decir, órganos cuyas decisiones son finales, en el sentido de que no existe un órgano superior que pueda revisarlas. La relevancia de la investigación sobre sesgos implícitos para el control de la decisión judicial individual está limitada, pero no por ello es irrelevante. Es limitada porque, por un lado, la decisión judicial se realiza en contextos deliberativos y no intuitivos,

más allá del papel que ciertamente le psicología del juez juega al momento de tomar la decisión. Por otro lado, porque, o bien existen criterios de corrección jurídica (para cualquier posición interpretativa, distinta del escepticismo radical, los jueces deciden dentro de cierto marco de criterios de corrección), o bien contamos ya con criterios explícitos para identificar decisiones discriminatorias. En esos casos, la corrección de la decisión estará fijada por estos criterios, independientes del sesgo.

De todas formas, las investigaciones sobre sesgos parecen adquirir relevancia cuando observamos las decisiones judiciales en su conjunto. Es decir, cuando nos preocupamos por el carácter sistemático de la discriminación. Desde este punto de vista, el problema es que decisiones individuales pueden aparecer correctas según criterios de corrección jurídica o criterios de control de la discriminación explícita, y sin embargo ser sesgadas si son consideradas dentro del conjunto de decisiones judiciales y se las compara con otras, o si se tiene en cuenta el resultado agregado de todas ellas. Se trata, sobre todo, de una exigencia de igualdad.

Ello permite advertir que, además del modo tradicional de controlar la actividad judicial mediante la posibilidad de recurrir, se vuelve también relevante un modo alternativo de control tendiente a evitar que los sesgos implícitos lleguen a producir el resultado discriminador, diseñando medidas para reducirlos o eliminarlos (incidiendo durante el proceso de selección de los jueces y, posteriormente, construyendo instancias de formación).

Por último, esta incidencia de los sesgos exige volver la mirada a los mecanismos previstos para el control de la actividad de los jueces individuales. Mecanismos tales como los consejos de la magistratura o jurados de enjuiciamiento. Sin embargo, este último punto exige abordar con detenimiento la cuestión de la responsabilidad de quienes deciden bajo la incidencia de sesgos. Al respecto, los sesgos implícitos representan un desafío para la teoría de la responsabilidad, puesto que no resulta del todo claro si está justificado responsabilizar a los sujetos por ellos. Por un lado, hay quienes sostienen una posición exoneradora, puesto que los sesgos implícitos son resultado de la educación, el contexto cultural y el estrato social de pertenencia, por lo que el agente no solo no puede controlarlos, sino que, además, suele no advertir su incidencia en el propio razonamiento. Si ello es así, las acciones sesgadas carecerían de dos rasgos que suelen asociarse con la responsabilidad: control y consciencia (Lawrence, 1987).

Por otro lado, distintas investigaciones muestran que, si bien el funcionamiento de los sesgos implícitos es inconsciente, de ello no se sigue que sea imposible para el agente advertirlos y, por lo tanto, controlarlos, por lo que resultaría, en cierto grado, responsable (Mavda, 2018). Para avanzar en esta disyuntiva suele proponerse extender la teoría de la responsabilidad por acciones emocionales al abordaje de la responsabilidad por acciones sesgadas. En efecto, el modo en que las emociones influyen en la conducta parece similar al modo en que lo hacen los sesgos implícitos. Así, en ciertas ocasiones las emociones resultan de una intensidad tal que desplazan, en la motivación del agente, el balance de razones a favor o en contra de la acción. Por ello, las acciones emocionales suelen ser consideradas como el resultado de factores causales que exceden el control del agente (Manrique, 2014). Pero esta será una tarea para futuros trabajos.

REFERENCIAS

Alchourrón, C. E. y Bulygin, E., 1971: *Normative Systems.* Wien-New York: Springer.

Appiah, K. A., 2000: "Stereotypes and the shaping of identity", *California Law Review,* 88 (1): 41-53.

Arena, F. J., 2016: "Los estereotipos normativos en la decisión judicial", *Revista de Derecho de la Universidad Austral de Chile,* 29 (1): 51-75.

Asensio, R. et al., 2010: *Discriminación de género en las decisiones judiciales: Justicia penal y violencia de género.* Buenos Aires: Defensoría General de la Nación.

Bagenstos, S., 2007: "Implicit bias, 'science', and antidiscrimination law", *Harvard Law & Policy Review,* 1: 477-493.

Caracciolo, R., 2009: "Discreción, respuesta correcta y función judicial", en *El derecho desde la filosofía. Ensayos.* Madrid: Centro de Estudios Constitucionales, 251-260.

Coloma, R., 2010: "El debate sobre los hechos en los procesos judiciales. ¿Qué inclina la balanza?", en Accatino, D. (ed.), *Formación y valoración de la prueba en el proceso penal.* Santiago de Chile: Abeledo-Perrot, 87-117.

Correll, J., et al., 2007: "Across the thin blue line: police officers and racial bias in the decision to shoot", *Journal of Personality and Social Psychology,* 92: 1006-1023.

De la Rosa Rodríguez, P. I. y Sandoval Navarro, V. D. 2016: "Los sesgos cognitivos y su influjo en la decisión judicial. Aportes de la psicología jurídica a los procesos penales de corte acusatorio", *Derecho Penal y Criminología,* 37 (102): 141-164.

Dworkin, R., 1984 [1978]: *Los derechos en serio.* Guastavino, M. (trad.). Barcelona: Ariel.

Dworkin, R., 2011: *Justice for Hedgehogs.* Cambridge-London: Belknap Press of Harvard University Press.

Fletcher, G. P., 1984: "Some unwise reflections about discretion", *Law and Contemporary Problems,* 47: 269-286.

Greenwald, A. y Hamilton Krieger, L., 2006: "Implicit bias: scientific foundations", *California Law Review,* 94 (4): 945-967.

Jolls, C. y Sunstein, C., 2006: "The law of implicit bias", *California Law Review,* 94: 969-996.

Kahneman, D., 2012: *Pensar rápido, pensar despacio.* 3.ª ed. Barcelona: Debate.

Lakoff, G., 1987: *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind.* Chicago-London: The University of Chicago Press.

Lawrence, C. R., 1987: "The Id, the Ego, and equal protection: reckoning with unconscious racism", *Stanford Law Review,* 39 (2): 317-388.

Manrique, M. L., 2014: *Dolo, estados mentales y responsabilidad penal.* México: Fontamara.

Mavda, A., 2018: "Implicit bias, moods, and moral responsibility", *Pacific Philosophical Quarterly,* 99 (S1): 53-78.

Mitchell, G. y Tetlock, P., 2006: "Antidiscrimination law and the perils of mindreading", *Ohio State Law Journal,* 67: 1023-1121.

Muñoz Aranguren, A., 2011: "La influencia de los sesgos cognitivos en las decisiones jurisdiccionales: el factor humano. Una aproximación", *InDret,* 2.

Oakes, P., Haslam, S. A. y Turner, J. C., 1994: *Stereotyping and Social Reality.* Oxford: Blackwell.

Peters, D. y Ceci, S., 1982: "Peer-review practices of psychological journals: the fate of published articles, submitted again", *The Behavioural and Brain Sciences,* 5 (2): 187-255.

Saba, R., 2009: "Igualdad, clases y clasificaciones: ¿qué es lo sospechoso de las categorías sospechosas?", en Gargarella, R. (ed.), *Teoría y crítica del derecho constitucional.* Buenos Aires: Abeledo Perrot, 695-742.

Saul, J., 2013: "Scepticism and implicit bias", *Disputatio,* 5 (37): 243-263.

Schauer, F., 2003: *Profiles, Probabilities and Stereotypes.* Cambridge, Mass.: Harvard University Press.

Steele, C. M., 2010: *Whistling Vivaldi. How Stereotypes Affect Us and what We Can Do.* New York: W.W. Norton & Co.

Tajfel, H., 1978: *Differentiation between Social Groups. Studies in the Social Psychology of Intergroup Relations.* London: Academic Press.

Von Wright, G. H., 1963: *Norm and Action.* London: Routledge and Kegan Paul. Hay traducción castellana de P. García Guerrero: *Norma y acción. Una investigación lógica.* Madrid: Tecnos, 1979.

Wistrich, A. y Rachlinski, J., 2018: "Implicit bias in judicial decision making. How it affects judgment and what judges can do about it", en Redfield, S. (ed.), *Enhancing Justice: Reducing Bias.* New York: American Bar Association.

Wistrich, A., Rachlinski, J. y Guthrie, C., 2015: "Heart versus head: do judges follow the law or follow their feelings", *Texas Law Review,* 93: 855-923.