Emociones y sesgos implícitos en el derecho: ¿fenómenos similares?(1)

Un abordaje desde las distintas teorías

Sofía Pezzano
Conicet; Universidad Nacional de Córdoba,
Facultad de Filosofía y Humanidades, Córdoba, Argentina.
pezzanosofia@gmail.com

Fecha de recepción: 15/12/2022 – Fecha de aceptación: 23/03/2023

Cómo citar este artículo: Pezzano, S. (2023). Emociones y sesgos implícitos en el derecho: ¿fenómenos similares? Un abordaje desde las distintas teorías. *Revista Perspectivas de las Ciencias Económicas y Jurídicas*. Vol. 13, N° 2 (julio-diciembre). Santa Rosa: FCEyJ (UNLPam); EdUNLPam; pp. 23-46. ISSN 2250-4087, e-ISSN 2445-8566. http://dx.doi.org/10.19137/perspectivas-2023-v13n2a02

Resumen: En el presente trabajo se desarrollan tres teorías de las emociones, y las críticas que se les realizan, para luego evaluar si es posible trazar similitudes analíticas entre las emociones y los estudios sobre sesgos implícitos, en términos de definición de ambos fenómenos y de consecuencias prácticas sobre la responsabilidad por las acciones emocionales o sesgadas.

Palabras clave: emociones; sesgos implícitos; responsabilidad.

(1) Este trabajo fue realizado en el marco del Proyecto de Investigación "Sesgos Implícitos en la decisión judicial. Responsabilidad institucional por discriminación", dirigido por Federico José Arena, radicado en y financiado por la Universidad Blas Pascal.

Agradezco profundamente las lecturas y comentarios de mis compañeros y compañeras de equipo: Paula Gastaldi, Francisco Manzanares, Luz Salomón, Victoria Gerbaldo y Carlos Martín Villanueva, con quienes compartimos varios años de debate y lecturas comunes. En especial, a Federico Arena, director del proyecto, por la dedicación en las diversas lecturas y la precisión en las observaciones. Además, agradezco la mirada técnica y especialista de María Laura Manrique; y la lectura atinada y cuidadosa de María Victoria Fernández.

Emotions and implicit biases in law: ¿similar phenomena? An approach from different theories

Abstract: This paper develops three theories of emotions, and their criticisms, and then looks at whether it is possible to draw analytical similarities between emotions and implicit bias studies, in terms of the definition of both phenomena and the practical consequences for responsibility for emotional or biased actions.

Key words: emotions; implicit biases; responsibility.

As emoções e os preconceitos implícitos no direito: fenómenos semelhantes? Uma abordagem a partir de diferentes teorias

Resumo: Este artigo desenvolve três teorias das emoções e as suas críticas e, em seguida, avalia se é possível estabelecer semelhanças analíticas entre as emoções e os estudos de enviesamento implícito, em termos da definição de ambos os fenómenos e das consequências práticas da responsabilização por acções emocionais ou enviesadas.

Palavras-chave: emoções; preconceitos implícitos; responsabilidade.

1. Introducción

¿Por qué las emociones le importan al derecho? En principio, porque parece que el derecho solo juzga acciones de los individuos que son **intencionales,** en el sentido de racionales, dirigidas o controladas por la razón, y las emociones, intuitivamente, parecen oponerse o ser un obstáculo para la razón. Sin embargo, al analizar las emociones, distintos autores y autoras critican esta visión y dan cuenta de sus aspectos racionales.

Por otro lado, en los últimos años han proliferado gran variedad de estudios sobre los sesgos cognitivos, particularmente sobre los sesgos llamados **implícitos** y su impacto en la supuesta imparcialidad y neutralidad que debe gobernar la toma de decisiones judiciales. Ejemplos de ello son los sesgos de anclaje, de confirmación y de retrospección.(2) Estos, al igual que las emociones, parecen estar fuera del ámbito de la racionalidad y de nuestro control. No obstante, cuando se realiza un análisis más complejo y profundo, las conclusiones sobre su aparente irracionalidad no son tan obvias como parecen. Por compartir esta particular apariencia de **irracionales** o **incontrolables,** y por poner en riesgo las exigencias de neutralidad e imparcialidad, los estudios sobre sesgos en el ámbito jurídico suelen apoyarse en algunas categorías analíticas aplicadas a las emociones, cuyas teorías se encuentran mucho más desarrolladas en los estudios jurídicos y filosóficos en general.

(2) El sesgo de anclaje es la tendencia a anclarse a los valores o datos dados inicialmente al momento de resolver un problema; el sesgo de confirmación es la tendencia a prestar atención a la información que confirma las creencias que ya tenemos, en lugar de la que la refuta; y el sesgo de retrospección es la tendencia a considerar que el resultado de un hecho que ya ocurrió era más previsible de lo que realmente lo es.

El presente trabajo introduce brevemente las distintas teorías de las emociones, y las críticas que se le realizan, para luego evaluar si es posible recurrir a ellas para abordar algunos aspectos de los sesgos implícitos. El objetivo es aportar claridad a los estudios sobre los sesgos y la responsabilidad, partiendo para ello del análisis de las teorías de las emociones por sus aparentes similitudes.

En la primera parte del trabajo se desarrollarán las principales teorías de las emociones, distinguiendo a grandes rasgos entre: teorías mecanicistas, teorías evaluativas y teorías mixtas o integradoras.(3) Luego, se definirá el concepto de **sesgo implícito** y se analizará su relación con las emociones a partir de la identificación de dos similitudes. Finalmente, se buscará determinar si estas similitudes identificadas son propiedades necesarias de las emociones y los sesgos, y si sirven para argumentar en relación a la responsabilidad por la portación de sesgos.

2. Tres teorías de las emociones

Cuando se habla de las emociones se hace referencia a un conjunto de hechos complejos que guardan entre sí algún "parecido de familia", pero que no comparten todas las características. González Lagier (2010) identifica seis elementos típicos o rasgos relevantes que suelen estar presentes en estos fenómenos diversos:

1) la creencia o evaluación;

2) el objeto intencional;

3) la reacción fisiológica;

4) la sensación;

5) la expresión de la emoción;

6) la disposición a la conducta.

(3) Hay otro conjunto de teorías que pueden englobarse bajo el rótulo de "teorías del giro afectivo". En este grupo se incluyen teorías que provienen de distintas tradiciones teóricas (feminismo, posestructuralismo, marxismo, etc.) que representan una ruptura con los supuestos de los que parten las formas tradicionales de estudiar las emociones en las ciencias sociales, basadas en la racionalización, para profundizar el estudio de las emociones y los afectos, y la emocionalización de la vida pública. En este trabajo no serán desarrolladas debido a que, si bien abordan la relación entre las emociones y el derecho, proponen una perspectiva completamente diferente a los tres enfoques que aquí se trabajan; es decir, no presentan una teoría alternativa, sino un cambio total de enfoque. Dentro de las autoras más relevantes de este conjunto de teorías se encuentra Sara Ahmed, con su libro *La política cultural de las emociones* (2015), el que recomiendo por su profundidad crítica. Allí, la autora critica las bases sobre las que se asientan tres teorías de las emociones que se desarrollan en el trabajo, relativizando la distinción entre la racionalidad y la irracionalidad, y mostrando que tiene consecuencias políticas relevantes. Muestra que los análisis de las emociones en términos de racionalidad, proporcionalidad, etc., necesitan de un parámetro moral sobre el cual analizar qué contenido de las emociones es correcto, cuánta intensidad es tolerable, lo que lleva a una selección de emociones buenas o correctas, por oposición a otras malas o incorrectas. Detrás de valoraciones presentadas como universales, sostiene, se esconde en realidad la moral dominante, y se excluyen otras formas de sentir y de expresarse que quedan relegadas al ámbito de la incorrección.

Los dos primeros elementos corresponden al aspecto evaluativo de las emociones. Las emociones tienen lugar a partir de una creencia determinada que puede ser de diversa índole: acerca de los hechos externos, de las emociones propias, de las emociones ajenas, de las motivaciones ajenas, de las creencias ajenas, etc. Esas creencias son de un tipo específico, ya que tienen un contenido evaluativo e implican un juicio acerca de un objeto o evento que le otorga una determinada significancia –negativa o positiva– en relación a un determinado objetivo, plan o deseo (creencia o evaluación).(4) Luego, las emociones son dirigidas o enfocadas a ese evento u objeto (objeto intencional).

Los tres elementos siguientes corresponden a la parte física o sensitiva. La emoción produce algunas reacciones en el cuerpo (reacción fisiológica), se **siente** en el cuerpo (sensación) y se expresa físicamente –cambios visibles en cuerpo, rostro, etc.– (expresión de la emoción). El último elemento tiene que ver con la acción. La persona que experimenta una emoción no siempre va a llevar a cabo determinada acción como consecuencia necesaria, pero el sujeto se siente impulsado a realizarla; que efectivamente la realice es una cuestión de hecho. La emoción implica una **tendencia** a la acción.

Resulta importante en este punto precisar el concepto de acción, ya que tendrá consecuencias en los posteriores análisis sobre la responsabilidad. Siguiendo a González Lagier (2010), se entenderá por acción a la producción voluntaria de un cambio en el mundo. El autor descompone este concepto en cinco elementos, a saber: una secuencia de movimientos corporales; una serie de cambios en el mundo; una conexión entre los movimientos y el cambio; una intención; y una interpretación o significado. En este sentido, señala:

> El agente se forma la intención de producir un determinado cambio; esta intención es llevada a la práctica por el agente a través de movimientos de su cuerpo. De alguna manera, la intención pone en marcha el cuerpo del agente, que realiza ciertos movimientos que sabe que se conectan (causalmente o de alguna otra forma) con el cambio pretendido, esto es, que son suficientes para producir el cambio, en circunstancias normales. (p. 80)

Las distintas teorías que se desarrollarán a continuación toman uno o varios de los seis rasgos relevantes que suelen estar presentes en las emociones –según González Lagier– y justifican su centralidad, sosteniendo que los demás son contingentes o no esenciales para la existencia de una emoción. Es decir, tienen distintas posiciones sobre cuáles son las propiedades necesarias y suficientes para que una emoción exista o, en otras palabras, desacuerdan sobre qué cuenta como emoción. Tradicionalmente, las emociones eran vistas como impulsos naturales o externos que llevaban a las personas a realizar acciones por fuera de su conciencia y/o control. Se vinculaban a lo irracional, se las relegaba al ámbito de la sensibilidad, de las meras sensaciones. Actualmente, en cambio, existe una tendencia a

(4) El desarrollo sobre este tipo específico de creencias con contenido evaluativo surge de González Lagier (2010, pp. 64--65), por lo que recomendamos su lectura para ampliar sobre esos conceptos. para ampliar sobre esos conceptos.

concebir a las emociones dentro del ámbito de la racionalidad. Existen, además, teorías mixtas que buscan integrar ambos aspectos. Dentro de cada una de estas tres grandes concepciones, se incluye un abanico de teorías muy amplio, cuya distinción resulta imposible realizar en el marco de este trabajo. A grandes rasgos, se identificarán con el rótulo de **mecanicistas, evaluativas** y **mixtas o integradoras**.

Estos tres grupos de teorías son relevantes en la medida en que la forma en que concebimos a las emociones influye en la evaluación que hacemos de las acciones emocionales, en la responsabilidad que se puede atribuir a la persona por su realización y en la posibilidad de educar y modificar las emociones.(5) Es decir, la discusión no se centra en el proceso físico o cognitivo que se produce cuando experimentamos una emoción, sino en las consecuencias que la forma de comprender ese proceso tiene en las acciones que realizamos y en el modo en que las evaluamos. Seguidamente, se desarrollarán las principales características de las teorías señaladas, para luego pasar a su relación con los sesgos cognitivos y sus consecuencias respecto de la responsabilidad.

2.1. Teorías mecanicistas

También conocidas como concepciones no racionales, ponen el acento en el conjunto de sensaciones que se experimentan al sentir una emoción, es decir, consideran que la esencia de las emociones es su aspecto fenomenológico –lo sentido– ya sea desde un punto de vista psicológico o fisiológico (González Lagier, 2010). Lo central de estas perspectivas es que creen que las emociones son sensaciones pasivas, irracionales y ajenas a nuestro control. Esta es la posición sostenida, entre otros, por autores como Descartes en *Las pasiones del alma* (1997), donde distingue el ámbito de las pasiones (a las que llama percepciones, emociones o sentimientos) del ámbito del pensamiento.

El rasgo de **pasividad** implica que se concibe a las emociones como una especie de fuerza ciega que experimentamos sin la intervención de nuestra voluntad, es decir, tenemos un rol pasivo frente a ellas. Además, se les atribuye la característica de ser **incontrolables,** en el sentido que no podemos elegir cuándo ni en qué medida experimentarlas. La **irracionalidad** tiene que ver con que no se encuentran dentro del ámbito de la razón, o más bien, que entorpecen los procesos de razonamiento. Todo esto implica que las emociones proporcionan una explicación causal de nuestras acciones, no basada en razones, y que no es posible modificar o reeducar la vida emocional de las personas, así como tampoco evaluarla moralmente (Kahan y Nussbaum, 1996).(6)

(5) Existe una diversidad de abordajes de las emociones y su influencia en la conducta y la responsabilidad. En particular, desde las neurociencias; véase, en este sentido, Vincent y Nadelhoffer (2013). Aquí se expondrán las tres grandes teorías filosóficas sobre las emociones, que funcionan como una especie de marco o paraguas que abarcan en su interior una variedad de perspectivas, a veces distintas entre sí.

(6) Es importante advertir que González Lagier, Kahan y Nussbaum no adscriben a las teorías mecanicistas, pero se citan aquí porque realizan una buena reconstrucción de esas teorías a los fines de realizar una crítica profunda.

Una virtud de esta concepción es su carácter parcialmente intuitivo, ya que captura una conexión entre las emociones y la pasividad, ese carácter externo que parecen tener, que muchas personas reconocen y expresan cuando hablan acerca de sus propias experiencias con distintas emociones. Por el contrario, en otros sentidos esta perspectiva resulta contraintuitiva, ya que, por ejemplo, en el derecho penal, algunas emociones sirven para excusar las conductas de las personas y otras no, e incluso, en algunos casos sirven para agravar la responsabilidad (Manrique, 2018).

Estas concepciones fueron, además, objeto de muchas otras críticas. En primer lugar, se sostiene que confunden las emociones con las sensaciones fisiológicas o psicológicas que se experimentan en el cuerpo en virtud de una emoción. Se alega, en contra de esta aserción, que es posible estar en cierto estado emocional y no percibir ninguna sensación en el cuerpo, situación que estas concepciones no pueden explicar (González Lagier, 2010).

En segundo lugar, se afirma que este grupo de teorías no puede explicar afirmaciones que suelen hacerse respecto de las emociones como, por ejemplo, que son razones para actuar (X realizó la acción Y **porque** sintió vergüenza), o que determinada emoción se encuentra justificada o injustificada (X **no debió** sentir ira en determinada situación) debido a que no identifican ninguna conexión racional entre la emoción y la acción. Es decir, al sostener que las emociones son fuerzas que empujan a las personas a actuar sin intervención de la razón y sin que medien juicios acerca del entorno, no permiten explicar el aspecto intencional de muchas de nuestras emociones (que se refieren a objetos, que implican creencias o evaluaciones acerca de ese objeto, etc.) (González Lagier, 2010). Veremos esto más ampliamente en los próximos apartados.

Además, nuestra forma de referirnos a las emociones no tiene que ver estrictamente con la descripción de una sensación física o psicológica. Es decir, no parece razonable describir una sensación física para manifestar que estamos experimentando una emoción; así, por ejemplo, en lugar de decir "sentí vergüenza", se puede afirmar "sentí un enrojecimiento de las mejillas".

En cuarto lugar, algunos autores y autoras destacan que las emociones implican ciertas creencias sobre el objeto al que se dirigen (Kahan y Nussbaum, 1996). Por ejemplo, X siente lástima porque cree que Y está sufriendo, pero si esa creencia cambia, porque Y le confirma que no está sufriendo, lo más probable es que la emoción cambie con ella, es decir, que X deje de sentir lástima. Los enfoques mecanicistas, al obviar el rol de las creencias en las emociones, fallan al explicar este punto. Muchas emociones, se señala, tienen un objeto **intencional,** lo que significa que su rol en la emoción depende de la interpretación que realiza la persona que experimenta la emoción, cuestión de la que los enfoques mecanicistas no pueden dar cuenta (Kahan y Nussbaum, 1996).

Por último, se afirma que hay muchas sensaciones similares que tienen lugar cuando experimentamos distintos tipos de emociones, por lo que describir los

estados psíquicos o físicos que se producen no sería suficiente para caracterizar una emoción determinada. Para lograr una explicación completa es necesario indagar en las creencias o pensamientos asociados a la emoción. Por ejemplo, puede haber una sensación de malestar físico similar en la vergüenza y en la compasión, o en la ira y el odio, pero son emociones diferenciables si pensamos más allá de lo que nos sucede en el cuerpo.

2.2. Teorías evaluativas

Los enfoques cognitivo-evaluativos son los que predominan entre las concepciones actuales. Sostienen, en términos generales, que "lo característico de las emociones su componente cognitivo, esto es, una creencia o una evaluación, con el que guardan algún tipo de conexión" (González Lagier, 2010, p. 45).

Marta Nussbaum (2008) defiende esta posición y sostiene que

> ... las emociones son una forma de juicio valorativo que atribuye a ciertas cosas y personas fuera del control del ser humano una gran importancia para el florecimiento del mismo. De esta manera, las emociones son efectivamente un reconocimiento de nuestras necesidades y de nuestra falta de autosuficiencia. (p. 44)

Es decir, que las emociones "siempre suponen la combinación del pensamiento sobre un objeto y el pensamiento sobre la relevancia o importancia de dicho objeto; en este sentido, encierran siempre una valoración o una evaluación" (p. 45).

La autora sostiene que las emociones tienen cuatro elementos constitutivos que determinan su identidad y permiten distinguir unas emociones de otras (Nussbaum, 2008). En primer lugar, tienen un objeto, son **acerca de** algo. En segundo lugar, ese objeto tiene carácter **intencional**, es decir, depende de la interpretación o percepción de la persona que experimenta una emoción. En tercer lugar, encarnan **creencias** acerca de ese objeto. Por último, esas creencias implican otorgar un **valor** a ese objeto, o una determinada importancia porque desempeña algún papel relevante en la vida de esa persona.

En otra publicación, Dan Kahan y Marta Nussbaum (1996) sostienen:

1) que las emociones encarnan creencias o puntos de vista que incluyen evaluaciones o valoraciones de la importancia de objetos y/o eventos;
2) que dichas evaluaciones pueden ser valoradas como apropiadas o inapropiadas; y
3) que las personas pueden (y a veces deberían) modificar sus emociones.

En relación al primer punto, afirman que las creencias o pensamientos no resultan contingentes, sino que son necesarios para identificar una emoción, forman parte de las emociones en sí mismas, las constituyen. Es decir, son propiedades **necesarias** de las emociones. Estas creencias no son de cualquier tipo, sino que son aquellas que implican una evaluación o valoración de determinados objetos o eventos como significativos. Valoramos el honor, por eso sentimos enojo

cuando alguien lo daña; valoramos nuestro cuerpo, por eso sentimos miedo cuando creemos que existe una amenaza de daño; valoramos a nuestros vínculos cercanos, por eso sentimos tristeza ante su pérdida, etc. (Kahan y Nussbaum, 1996).

Con respecto a la segunda característica, las evaluaciones que hacemos de los objetos o eventos, a su vez, pueden ser sometidas a evaluación. Es posible sostener que una persona que tiene mucho miedo frente al posible robo de su celular está haciendo una evaluación inapropiada de ese objeto –el celular– y de ese evento –el posible robo–, o como mínimo, desmedida.

Entonces, hasta aquí es posible reconocer distintos tipos de evaluaciones relacionadas a las emociones. La primera de ellas es constitutiva de la emoción, y es la evaluación que se hace de un objeto o evento como significativo. A partir de esa evaluación que llamamos **constitutiva,** es posible realizar otra, posterior a la emoción, que es la evaluación del valor que se le da a ese objeto o evento. Allí es posible concluir que la evaluación constitutiva resulta apropiada, inapropiada, razonable, irrazonable, etc. Esta evaluación es **externa** a la emoción. Puede tener como resultado que una emoción resulte inapropiada porque le dio un excesivo valor al objeto o evento, por ejemplo, sentir mucho miedo por la posibilidad del robo de un celular (Kahan y Nussbaum, 1996).

En este punto, es posible realizar otras distinciones relacionadas no ya con la emoción en sí, sino con las acciones realizadas por las personas en virtud de experimentar determinada emoción. Una primera evaluación la realiza la persona que experimenta la emoción: analiza qué acción realizar en virtud de la emoción. Es una evaluación **interna.** Cabe aclarar que **no siempre** que se experimenta una emoción, se actúa en consecuencia. Sin embargo, aunque finalmente se decida no actuar, la evaluación sigue presente: lo que es contingente es la acción, no la evaluación (el resultado de esta puede ser decidir **no actuar***)*. Supongamos que es probable que le roben el celular a X (valoración del objeto celular como importante y del evento robo como probable) y que siente mucho miedo de que ese celular sea robado (emoción que podemos valorar como excesiva, inapropiada). Frente a esto, la persona que siente la emoción evalúa de manera interna qué acción llevar a cabo. Por ejemplo, podría decidir no usar el celular en la calle o, de manera más extrema aún, según la intensidad de la emoción, podría decidir no llevar consigo el celular cuando salga a la calle.

Un segundo tipo de evaluación relacionado a la acción es aquella que la valora por su razonabilidad en virtud de la proporcionalidad con el objeto o evento valorado. Esta evaluación es **externa.** Por ejemplo, no salir nunca de su casa por miedo a que le roben el celular es irrazonable. En este punto, la evaluación puede dar como resultado que la acción concuerda con la emoción, pero es irrazonable en función del objeto: se está teniendo un cuidado excesivo (no salir de la casa) que no es proporcional al valor del objeto (celular). Este segundo punto tiene que ver con el resultado de la evaluación externa señalada

anteriormente, que determina si el objeto está correcta o incorrectamente valorado como significativo.

Ahora bien, si X está utilizando el celular, se le acerca otra persona y X la mata porque tenía mucho miedo de que le robe el celular, ese curso de acción, aunque se condiga –por su intensidad– con la emoción, no se encuentra justificado porque realizó una valoración incorrecta del objeto y del evento (le dio un excesivo valor al celular, mayor al de la vida de otra persona). Aunque la otra persona haya tenido la intención de robarle el celular, aquí no aceptaríamos que la emoción justifica la acción, ya que no era razonable tener tanto miedo en esa situación, ni apropiado actuar de esa manera.

La evaluación externa de la acción, entonces, se relaciona con el valor que el sujeto le dio inicialmente al objeto o evento que da lugar a la emoción. Una acción puede ser proporcional a la emoción, pero no al objeto que da lugar a la emoción. En el primer sentido, lo que se valora es si se actúa **en la medida** de la emoción sentida. En el segundo, lo que se valora es si la acción se encuentra justificada en relación a la valoración racional del objeto significativo; aquí lo relevante es si esa evaluación del objeto incluida en la emoción era razonable o no.

Pensemos en otra situación: X (mujer) está a punto de ser asesinada por Y (varón), siente mucho miedo y lo hiere con un cuchillo. La valoración del objeto –su vida– es correcta, la emoción sentida en virtud de esa valoración –mucho miedo– es apropiada y proporcional. La acción realizada se condice con el tipo de emoción, es decir, está justificada en relación a la emoción (la emoción fue expresada apropiadamente en la acción) y en relación a la valoración del objeto (su vida). En este caso, en todos los aspectos, X hizo lo que haría cualquier persona razonable si se encontrara en su misma situación.

En conclusión, las acciones emocionales pueden implicar un entramado de cuatro tipos de evaluaciones:

1) la del objeto y el evento en sí mismo;

2) la de la emoción que se sigue de esa valoración;

3) la de la acción a realizar;

4) la de la acción realizada.

La primera es interna y constitutiva de la emoción, es decir, es la que origina la emoción. La segunda es externa y evalúa si esa valoración del objeto o evento es apropiada o inapropiada. La tercera es interna porque tiene que ver con cómo actuar en virtud de la emoción sentida, o con la decisión de no tomar ningún curso de acción. La cuarta es también externa, porque evalúa la relación entre acción, emoción y objeto. Los sujetos evaluadores no son necesariamente los mismos en todos los casos: la evaluación 1 y 3 es realizada por quien experimenta la emoción y por quien realiza/no realiza la acción. Las otras dos pueden ser llevadas a cabo por esa misma persona o por otra diferente.

Cuando estas teorías se refieren a que una emoción es irracional, no estamos diciendo que una fuerza externa se ha apoderado de la persona que experimentó esa emoción. Por el contrario, esta afirmación puede referirse a cuatro situaciones diferentes:

1) que se ha hecho una evaluación incorrecta de un determinado objeto o evento;
2) que la emoción sentida no es proporcional a la importancia de ese objeto o evento;
3) que la acción realizada no encuentra relación con la emoción; o
4) que la acción realizada no encuentra relación con la significancia del objeto o evento (Duff, 2015).

Anteriormente se mencionó que Nussbaum y Kahan (1996) reconocen tres aspectos de las emociones. Ya desarrollados los primeros dos (que las emociones implican creencias y evaluaciones), es momento de pasar al tercero: la posibilidad de modificar nuestras emociones y de educarnos emocionalmente. Esto se deriva directamente del hecho de que las emociones son consideradas racionales y, por ende, es posible, a través de la razón, modificarlas. En este punto se abren nuevas discusiones, que no abordaremos en este trabajo, sobre cómo es posible modificar esas emociones, y si la responsabilidad por modificar las emociones equivocadas es individual o es un problema que debe abordarse de manera estructural.

Lo que el enfoque evaluativo quiere resaltar es que constantemente estamos realizando evaluaciones de este tipo cuando nos encontramos frente a emociones y acciones emocionales, y esto no tendría sentido si considerásemos que las emociones son fuerzas externas e irracionales que nos controlan sin que podamos intervenir.

En resumen, las concepciones evaluativas sostienen que las emociones implican juicios acerca de ciertos objetos o situaciones (llamados **intencionales)**. Las emociones pueden ser objeto de evaluación, es decir, considerarse justificadas o injustificadas, adecuadas o inadecuadas, lo que se vincula directamente con que el juicio que se realiza sobre un determinado objeto o situación sea justificado, adecuado, etc. Además, las emociones pueden ser razones para la acción (explicaciones no meramente causales de las acciones), y esas acciones también pueden ser evaluadas. Por último, al ser racionales, pueden ser controladas y, por ende, reeducadas (González Lagier, 2010).

Una de las críticas que se le suele hacer a este tipo de enfoques es que son reduccionistas al dejar de lado una parte relevante del fenómeno, que es que las emociones se encuentran asociadas a sensaciones de placer o de dolor. Aristóteles ya había reconocido este aspecto. Las emociones se sienten, y esto es lo que nos permite sostener que hay emociones positivas y negativas (en relación al placer o al dolor que percibe en el cuerpo la persona que experimenta una

emoción determinada). Por ejemplo, cuando sentimos ira o vergüenza, nuestro cuerpo experimenta una serie de sensaciones dolorosas, como el enrojecimiento de la piel, calor interno, sensación de que se nos cierra el pecho, etc. Esta característica de las emociones es lo que hace que las veamos como algo externo, que no podemos controlar y que simplemente nos sucede, sin que tengamos un rol activo en ello (González Lagier, 2010).

Por otra parte, se afirma que la perspectiva evaluativa es excesivamente racionalista al definir a las emociones como un conjunto de creencias y deseos, lo que lleva a construir una imagen distorsionada de las mismas (González Lagier, 2010).

Por último, algunas autoras sostienen que no permite percibir la relevancia de la obstaculización del juicio que experimenta una persona al experimentar algunas emociones fuertes provocadas por la presión de circunstancias externas (Uniacke, 2007), aunque este aspecto resulta muy controvertido.

2.3. Teorías mixtas o integradoras

González Lagier (2010) sostiene que las teorías presentadas anteriormente lo que hacen es identificar alguno de los elementos presentes en las emociones y otorgarle un estatus esencial (sensaciones, creencias, evaluaciones, etc.). Para el autor, esa estrategia no es satisfactoria porque o bien ese elemento identificado no es **necesario** –no aparece en todos los casos que nos encontramos frente a una emoción– o no es **suficiente** –existen otros elementos o aspectos que las definen–.

Frente a este escenario, surgen diversas teorías que buscan combinar las características centrales de los enfoques mecanicistas y evaluativos. Algunas de ellas buscan dar una definición no reduccionista de las emociones, uniendo los elementos identificados por aquellos enfoques. Otras afirman que el fenómeno de las emociones es tan amplio e incluye situaciones tan diversas que no es posible brindar un concepto acabado en términos de propiedades necesarias y suficientes, o que no es recomendable, ya que resultaría subincluyente, entonces ofrecen un conjunto de características que están presentes en algunos casos, pero no en todos. Esta posición es la que sostienen autores como Suzanne Uniacke, Laura Manrique, John Elster y González Lagier. Me enfocaré en la perspectiva de este último autor.

González Lagier (2010) parte del supuesto de que tanto los aspectos resaltados por las teorías mecanicistas como los aspectos cognitivo-evaluativos forman parte de las emociones, de manera tal que ninguno puede ser dejado de lado. Las emociones son duales: son tanto razones para la acción como causas de la misma; apoyan a la razón, pero también la limitan. Esta dualidad repercute en la posibilidad de controlarlas y modificarlas, y en la atribución de responsabilidad por acciones emocionales. Entonces, desde esta perspectiva, ¿es posible controlar nuestras emociones? ¿somos responsables por ellas? ¿en qué medida?

Las emociones tienen dos papeles distintos en la génesis de una acción: contribuyen a formar la intención y limitan causalmente las alternativas de acción. En el primer punto se identifica el papel de la emoción en el entramado de las razones para la acción, mientras que en el segundo se identifica a la emoción como una causa mecánica de la acción. Como las emociones son complejas y se componen tanto de elementos racionales como mecanicistas, cuando estamos frente a una acción que tiene origen en una emoción, son necesarios los dos tipos de explicaciones para dar cuenta del fenómeno (González Lagier, 2010).

Teniendo en cuenta las distinciones establecidas, parece obvio que la conclusión es que las acciones realizadas bajo el influjo de una emoción están **parcialmente** dentro del control del agente. En nuestra tradición jurídica, principalmente en el ámbito del derecho penal, las emociones se han utilizado como razón tanto para atenuar la responsabilidad por determinadas acciones como para agravarla. Es decir, las emociones modifican la responsabilidad, pero de diferentes maneras. Para poder explicar ambos efectos, según González Lagier, es necesario acudir a una teoría dual de las emociones, debido a que las concepciones mecanicistas y cognitivo-evaluativas solo pueden brindar explicaciones parciales. Así, sostiene que "las emociones son razones para la acción, pero, en otro nivel, son también causas de la misma que disminuyen el margen de libertad" (González Lagier, 2010, p. 148). Esto último lo pueden hacer de manera más o menos extrema en función de su **intensidad**:

> Las emociones no excluyen la elección, incluso la posibilitan, pero cuanto más intensas son, más reducen el campo de actuación de nuestra razón. Nuestra razón necesita de las emociones, pero llega un momento en que se basta por sí sola. Si la emoción va más allá, su ayuda se vuelve entorpecimiento. Las emociones son ambivalentes. Son como un foco que ilumina cierto aspecto del mundo, dejando a la penumbra el resto. (p. 149)

La característica de la **intensidad** es la que explica por qué en algunas situaciones se le disminuye la responsabilidad a una persona por una acción realizada bajo el influjo de una emoción. Por ejemplo, en los casos de miedo insuperable, emoción violenta, entre otros. A este aspecto debemos sumarle el análisis del **contenido** de la emoción. Una emoción puede ser particularmente intensa, pero inapropiada, por ejemplo, por basarse en consideraciones misóginas o racistas. Este contenido también tiene efecto en la responsabilidad, y explica por qué en algunos casos se agrava la responsabilidad. Por ejemplo, en los crímenes de odio.

Estas dos características pueden combinarse y resultar en cuatro casos posibles:

1) emociones intensas basadas en creencias justificadas;

2) emociones intensas basadas en creencias injustificadas;

3) emociones frías basadas en creencias justificadas; y

4) emociones frías basadas en creencias injustificadas (p. 150).

El cuarto caso claramente deriva en un agravamiento de la responsabilidad porque no se cumple ninguno de los dos requisitos para su atenuación (intensidad y contenido apropiado). El primer caso es un caso claro de atenuación de responsabilidad porque ambos requisitos están presentes. En los otros dos casos es donde se requiere un mayor cuidado, ya que, para González Lagier, en ambos casos, de mínima, no corresponde atenuar la responsabilidad.(7)

Adjudicar a una persona responsabilidad por sus acciones emocionales depende, según el autor, de que puedan ser evaluadas como racionales o irracionales;(8) de que tengan capacidad para motivar acciones; y de que podamos controlar las emociones y las acciones al menos parcialmente (González Lagier, 2009). González Lagier considera, además, que las emociones satisfacen estos tres requisitos. Con respecto al último, señala que las principales vías de control de las emociones son tres: la revisión de nuestras creencias; la manipulación del contexto en el que se puede originar una determinada emoción; y la revisión de nuestros deseos o fines (González Lagier, 2010). No se profundizará, sin embargo, en este punto.

A modo de cierre, es posible afirmar que, en términos generales, esta teoría se muestra superadora de las críticas realizadas a las teorías mecanicistas y evaluativas. Por un lado, no identifica a las emociones como fuerzas externas, irracionales y fuera del control de los sujetos. Se señaló anteriormente que esta forma de definirlas es contraintuitiva en varios sentidos y no permite dar cuenta de otros aspectos relevantes ni puede explicar las formas en las que solemos hablar de las emociones. Por otro lado, no excluye los aspectos físicos-sensitivos, que no dejan de existir porque sostengamos que las emociones son fenómenos racionales y que se encuentran –aunque sea parcialmente– dentro de la esfera de nuestro control. Además, brinda razones contundentes para atribuir responsabilidad por acciones emocionales, y esta perspectiva se adecúa en gran medida al tratamiento de nuestras emociones que hace el derecho. Un claro ejemplo de ello es nuestro Código Penal (González Lagier, 2009).

(7) De manera general, el autor considera que para que exista una atenuación de responsabilidad, es condición necesaria, aunque no suficiente, la existencia de intensidad. Por su parte, para que exista un agravamiento de la responsabilidad, es condición necesaria, aunque no suficiente, la existencia de una **creencia inapropiada.** Para profundizar este punto, véase González Lagier (2009).

(8) González Lagier (2009) sostiene que "un defecto en la justificación de la creencia que suscita la emoción hará que la emoción correspondiente sea irracional o no esté epistémicamente justificada. Así, nos podemos encontrar con emociones irracionales por basarse en creencias dogmáticas o sin evidencia a su favor, emociones irracionales por basarse en una creencia derivada a partir de creencias a su vez injustificadas, emociones irracionales por basarse en una inferencia inadecuada" (p. 446). Véase, para ejemplos de cada una, González Lagier (2009, p. 446). Continúa el autor: "... un segundo tipo de irracionalidad de la emoción, también vinculado con las creencias que las generan, es el que se da por falta de correspondencia entre el tipo de creencia y el tipo de emoción", y agrega después: "Una variante del anterior tipo de irracionalidad emocional es el de las emociones excesivas o, por el contrario, el de las emociones insuficientes" (p. 446).

3. Relación entre emociones y sesgos

Las emociones se relacionan con los sesgos implícitos porque parecen tener algunos parecidos de familia y, por ende, consecuencias similares (Madva, 2017).

Existen muchas discusiones teóricas alrededor del concepto de sesgos implícitos, y los autores y autoras lejos están de llegar a un acuerdo. También se encuentran bajo discusión las credenciales científicas de las investigaciones sobre sesgos, y en particular de los tests que suelen realizarse para comprobar si una persona porta un determinado sesgo o no.(9) En lo que resulta relevante para este trabajo, los debates teóricos se centran, por un lado, en intentar delimitar el significado de **sesgos** y, por el otro, en identificar qué implica que sean **implícitos**.

De la Rosa y Sandoval (2016) sostienen que los sesgos "permiten 'reducir las tareas complejas de asignar probabilidad y predecir valores a operaciones de juicio más simples'" (p. 148). Un ejemplo de ello es el sesgo de anclaje, que es la tendencia a atenerse o "anclarse" a los datos aportados inicialmente a la hora de abordar un problema. Este sesgo es común en los juicios de daños y perjuicios, cuando el demandante solicita un monto determinado como indemnización, y el juez o jueza otorgan esa cantidad o un monto apenas menor, sin realizar los cálculos correspondientes, porque el valor inicialmente dado funciona como "ancla".(10) Como vimos en este caso, estos procesos pueden tener como resultado errores en el razonamiento. Como puede notarse, estos autores consideran que los sesgos implican un conjunto de procesos cognitivos que las personas llevan a cabo. Son especies de atajos en el razonamiento. En este sentido, el concepto es descriptivo.

Sin embargo, otros autores asignan al concepto una connotación negativa, es decir, para ellos se trata de un concepto normativo (Stafford *et al.*, 2018). Así, los sesgos son vinculados a evaluaciones distorsionadas y negativas de determinados sujetos o grupos. Estas concepciones ven al resultado –error en el razonamiento– del proceso cognitivo como parte del proceso en sí mismo. En este sentido, Frankish (2016) sostiene que "una persona sesgada es aquella que se dispone a juzgar a otras de acuerdo a concepciones estereotipadas de su grupo social (etnia, género, clase, y demás) en vez de por sus talentos individuales" (p. 2).(11)

Ahora bien, ¿qué significa que estos procesos cognitivos, además, sean **implícitos**? Stafford *et al.* (2018) señalan que, para muchos autores y autoras, implícito es equivalente a **inconsciente,** es decir, sesgos que las personas no saben que tienen. Sin embargo, distintos estudios empíricos han demostrado evidencia en

(9) Para profundizar en estas discusiones, véase Mitchell y Tetlock (2006).

(10) Véase Canale *et al.* (2021) para un desarrollo más detallado de este ejemplo y del sesgo de anclaje.

(11) La traducción me pertenece.

contrario.[12] Otros sostienen que implícito significa que **no pueden controlarse de manera directa** por la persona que porta el sesgo, lo que implica que es posible alguna forma indirecta de control o mitigación de sus efectos. En contra de esta posición, se sostiene que no puede ser un aspecto suficiente para caracterizar este tipo de procesos, porque existen muchos otros procesos cognitivos y estados –que incluso pueden caracterizarse como explícitos– que están más allá de nuestro control directo –por ejemplo, las emociones, las creencias, etc.–. Otra forma de definir lo **implícito** es entenderlo como **disonante** en relación a otras creencias que el agente sostiene explícitamente. Esta es la posición que sostiene Frankish (2016). Sin embargo, Stafford *et al.* (2018) sostienen que tampoco se terminan de captar todos los fenómenos a los que nos referimos cuando hablamos de sesgos implícitos porque muchas veces existen aun siendo compatibles con las actitudes explícitas de las personas.

Es notable cómo la manera en que definimos el concepto de sesgos implícitos influye en los tipos de fenómenos que captamos y en las consecuencias normativas que les asignamos. El concepto de sesgos utilizado por De la Rosa y Sandoval, por ejemplo, distingue los procesos cognitivos de los resultados que puedan tener en el razonamiento de las personas. Ese concepto ya incluye como requisito necesario que ese proceso mental sea inconsciente. Es decir, que para estos autores no existiría la posibilidad de la existencia de sesgos no implícitos.

Una distinción que resulta muy útil para entender a qué nos referimos cuando hablamos de sesgos es de los conceptos de prejuicio y discriminación. En este sentido, Mitchell y Tetlock (2006) sostienen que

> Un "sesgo" se refiere a la variación sistemática de las tendencias de juicio provocada por algún atributo o propiedad de un estímulo, como la pertenencia a un grupo particular (...) El sesgo puede ser implícito, en cuyo caso la gente no reconoce la influencia de un estímulo en su juicio en el momento de su funcionamiento, o explícito, en cuyo caso la gente reconoce la influencia. Un sesgo cognitivo implícito, así definido, opera automáticamente, porque se produce en el momento del juicio más allá de la conciencia consciente o el control intencional. Sin embargo, las condiciones previas deben ser las adecuadas para que un sesgo implícito se active automáticamente y luego, a su vez, influya en el comportamiento exterior. Y un sesgo implícito puede hacerse explícito de repente si el contexto social alerta a la gente sobre la dirección y la magnitud del sesgo, haciendo así posible la autocorrección. (Mitchell y Tetlock, 2006, p. 1035)[13]

Por su parte, los autores sostienen que **prejuicio** "se refiere a una respuesta afectiva o evaluativa sistemática a un grupo social y sus miembros (...) es una actitud especial reservada para grupos, y las actitudes prejuiciadas pueden ser también implícitas o explícitas de la misma manera" (Mitchell y Tetlock, 2006, p. 1036).[14]

(12) Para ampliar sobre este punto, véase Holroyd, Scaife y Stafford (2017).

(13) La traducción me pertenece.

(14) La traducción me pertenece.

Por último,

> La "discriminación" se refiere a las consecuencias conductuales de un sesgo de grupo (típicamente en forma de un estereotipo de grupo) o un prejuicio hacia un grupo en particular. Bajo esta definición psicológica expansiva, cualquier consecuencia conductual de un sesgo o prejuicio de grupo cuenta como discriminación. (Mitchell y Tetlock, 2006, p. 1036)[15]

Es decir, que el sesgo es cognitivo y el prejuicio es emotivo o actitudinal; ambos pueden ser implícitos o explícitos. Son implícitos cuando operan automáticamente en el momento de realizar un juicio o una valoración sobre determinada persona o grupo. Si se traducen en conductas, se trata de discriminación. Este concepto de discriminación es amplio porque no distingue si el contenido es positivo o negativo. Estas disquisiciones resultan útiles en la medida en que sirven para separar a los sesgos de otros fenómenos asociados o similares.

¿Es posible trazar un paralelo entre las concepciones de emociones desarrolladas y las distintas maneras de entender a los sesgos implícitos? ¿cuáles son –si las hay– las similitudes analíticas que permitirían trazar este paralelo y otorgarles el mismo tratamiento? La asunción de que las emociones se parecen en algunos sentidos a los sesgos y que, por lo tanto, comparten algunos rasgos esenciales que permiten que se realicen preguntas similares acerca de la responsabilidad y posibilidad de modificación, se basa en que los autores y autoras identifican –al menos– dos características que parecen estar presentes en ambos fenómenos: la **apariencia** de inconciencia y la **apariencia** de falta de control o de control reducido.

Resulta difícil realizar un paralelo entre fenómenos sin tomar posición sobre alguno de los conceptos de **emoción** y **sesgo**. A los fines de realizar esta comparación, se tomará la teoría mixta de las emociones de González Lagier. En el caso de los sesgos, se tomará la teoría de Madva –que se desarrollará en el próximo apartado–.

Antes de entrar de lleno en la comparación, cabe realizar una aclaración terminológica. Madva, técnicamente, compara los sesgos con los **estados de ánimo** *(moods)* y no con las emociones, y señala que deberíamos dar el mismo tratamiento a los sesgos que a los estados de ánimo en lo que hace a la responsabilidad. Ahora bien, ¿cuál es la relación entre las emociones y los estados de ánimo? No hay acuerdo al respecto. Algunos/as sostienen que son fenómenos distintos –aunque relacionados–, otros/as afirman que tienen una relación género-especie, y para otros/as que no está claro en qué se diferencian. González Lagier (2010), adscribiendo a esta última posición, sostiene que

> ... [a]lgunos autores distinguen entre emociones y estados de ánimo. Estos últimos serían semejantes a las primeras, pero tendrían una duración más prolongada, carecerían de un objeto definido y –a veces– presentarían una activación fisiológica menos intensa (por ejemplo, la melancolía prolongada, la ansiedad continuada, el

(15) La traducción me pertenece.

> enamoramiento duradero, etc.). Dada la vaguedad de estos criterios, no resulta fácil distinguir entre emociones y estados de ánimo. Además, muchos estados de ánimo y muchas emociones tienen el mismo nombre. E, incluso, algunos estados de ánimo pueden considerarse "huellas" dejadas por emociones intensas. Dado que no hemos tomado a los elementos de las emociones como condiciones necesarias y suficientes (esto es, algunas emociones pueden carecer de uno o varios de estos elementos, o tenerlos en distinto grado), podemos considerar los estados de ánimo como emociones debilitadas, aunque persistentes. (pp. 76-77)

Madva, en su artículo, se basa en la diferencia entre emociones y estados de ánimo realizada por Beedie *et al.* (2005).[16] Estos autores sostienen que se trata de fenómenos distintos, aunque relacionados. Algunos aspectos en los que se distinguen son:

1) causa: las emociones son originadas en algún evento específico localizado temporalmente, mientras que los estados de ánimo se originan como consecuencia de concatenaciones de pequeños incidentes, condiciones persistentes del ambiente y/o procesos internos metabólicos o cognitivos;
2) duración: las emociones suceden en un instante, mientras que los estados de ánimo son más duraderos;
3) control: las emociones son más difíciles de controlar que los estados de ánimo;
4) experiencia: las emociones se relacionan con los sentimientos, y los estados de ánimo con los pensamientos;
5) consecuencias: las emociones no nublan el juicio, mientras que los estados de ánimo sí lo hacen;
6) visibilidad: las emociones son públicas y los estados de ánimo personales;
7) intencionalidad: las emociones siempre se dirigen a algún objeto o evento, mientras que los estados de ánimo no siempre;
8) intensidad: las emociones son más intensas que los estados de ánimo;
9) fisiología: las emociones suelen estar relacionadas con procesos fisiológicos, mientras que los estados de ánimo no;
10) tiempo: las emociones suelen tener un momento claro en el tiempo en que se originan, que los estados de ánimo no necesariamente tienen;
11) estabilidad: las emociones tienden a ser más constantes que los estados de ánimo;
12) conciencia de la causa: en las emociones, las personas pueden ser conscientes más fácilmente de la causa que las origina que en el caso de los estados de ánimo;

(16) Véase nota 19 en Madva (2017).

13) claridad: las emociones son más claramente identificables y explicables que los estados de ánimo; y

14) función: las emociones sesgan la conducta, y los estados de ánimo sesgan la cognición (pp. 864-870).

En el trabajo de Beedie *et al.* (2005), sin embargo, puede observarse que no hay acuerdo sobre esas características (de hecho, lo muestran en términos porcentuales), aunque la mayoría de los/as autores/as relevados/as y de las personas no académicas entrevistadas coinciden en que esas son las diferencias. Todas esas características de las emociones, además, son debatibles. Vimos anteriormente que, por ejemplo, no toda emoción tiene como consecuencia una acción, que algunas emociones son más intensas que otras, que muchas emociones pueden ser controladas, que muchas veces se tiene una emoción sin ninguna **sensación** que la acompañe, etc.

La cuestión es que, como surge de la cita precedente de González Lagier, las emociones son fenómenos de los que no pueden predicarse condiciones necesarias y suficientes, lo que hace que en muchos casos varias de las características idecas señaladas por Beedie *et al.* (2005) pueden ser características de las emociones, pero también pueden darse casos en que algunas no se den y aun así podamos seguir hablando de emociones.

Ahora bien, Madva (2017) identifica dos características de los estados de ánimo para compararlos con los sesgos, que también son adscribibles a las emociones –siempre con la salvedad de que no se trata de condiciones necesarias ni suficientes para que exista una emoción–: la conciencia (limitada) y el control (limitado).

Por estas razones, las consideraciones que hace Madva sobre los estados de ánimo pueden ser trasladadas a las emociones tal como son entendidas en el presente trabajo. Hecha esta aclaración, corresponde analizar la posición de Madva.

3.1. Rasgos similares: conciencia y control

La comparación que realiza Madva (2017) entre emociones y sesgos implícitos tiene la finalidad de argumentar a favor de la responsabilidad gradual en caso de portación de sesgos, por **analogía** con lo que –según sostiene– es aceptado en caso de responsabilidad por emociones. Sin embargo, acude para ello a una visión intuitiva y simplificada de lo que son las emociones. No da un concepto ni identifica propiedades relevantes, sino que desarrolla las similitudes a través de un ejemplo: Gertie siente enojo, puede no saber sus causas o incluso no notarlo, pero ninguna persona pensaría que no es responsable por las acciones que realiza (como interrumpir a una amiga en una conversación) solo porque se encuentra enojada; sin embargo, estar enojada sí sirve como una explicación parcial de los comportamientos y puede afectar de algún modo la responsabilidad por sus actos. Las emociones mitigan la responsabilidad y esto, sostiene Madva,

se prueba por el modo en que ofrecemos y aceptamos disculpas por comportamientos influenciados por emociones. Luego sostiene que, sin embargo, es necesario tener en cuenta el tipo de acción o comportamiento que la emoción pretende mitigar (cuanto más graves son las consecuencias de las acciones, menos efectos mitigadores va a tener la emoción).

Con respecto a la conciencia *(awareness)* de la portación de un sesgo implícito, Madva ofrece una serie de argumentos empíricos con los que pretende demostrar que las personas sí se dan cuenta de que portan un sesgo. Así, sostiene que el contenido de los sesgos está disponible a la experiencia consciente, aunque en muchos casos no es objeto de atención explícita:

> Las pruebas se entienden mejor a la luz de una distinción familiar entre el contenido de la fenomenología propia (es decir, lo que se experimenta) y el contenido de su atención focal. Es decir, pueden sentirse sin que se noten, al igual que una persona puede estar de mal humor o alegre sin darse cuenta. (p. 8)

El autor afirma que uno de los efectos de las emociones es el de **sesgar** la cognición, llevando a los individuos a prestar atención, ignorar, malinterpretar o percibir distinto algunas características de su entorno. De nuevo, Madva apela a que las emociones –al igual que los sesgos– tienen esta característica de que pueden pasar desapercibidas, en el sentido de que quien las porta no nota que se encuentra bajo el efecto de una emoción y realiza acciones en virtud de ello. Sin embargo, sostiene que esto no excusa por completo de la responsabilidad porque hay algún tipo de percepción tácita o potencial que hace a quienes actúan al menos parcialmente responsables de sus acciones. A raíz de esta disponibilidad a la experiencia consciente, pero de su posibilidad de no ser notados, Madva concluye que la conciencia en los sesgos, al igual que en las emociones, se da **en grados**.

En relación a la posibilidad de control, el autor sostiene que cuando actuamos de manera sesgada, hacer referencia al sesgo como forma de justificar ese accionar puede tener consecuencias respecto a la responsabilidad que se me atribuye, porque en algún punto el sesgo no me permitió **controlar** mi acción. Apela a las emociones, nuevamente, para explicar su similitud: si X dice que no saludó a Y cuando se la cruzó en la calle **porque** estaba enojado, es en algún punto una justificación. Pero también Y podría decirle a X que intente manejar su humor, o que aprenda a quién dirigirlo. Es decir, ambas formas de referirnos a los sesgos y a las emociones son sensatas y las utilizamos en la vida diaria. Es por ello que Madva (2017) sostiene que su teoría de control gradual es la más apropiada para captar los fenómenos.

Para Madva, es posible controlar tanto los sesgos implícitos como las emociones, pero es cierto que resulta difícil hacerlo. El control, sostiene, también es una cuestión de grados. Distingue tres maneras en las que podemos pensar el control: local, indirecto y a largo plazo. La primera de ellas tiene que ver con la posibilidad de dar un paso atrás en el momento en que opera el sesgo, es decir, es un control reflejo. El segundo tipo de control tiene que ver con la posibilidad

de manipular el contexto en el que suele generarse el sesgo. El tercero apunta más a la educación para lograr eliminarlos a largo plazo. Estos puntos son muy similares a los señalados por González Lagier cuando se abordó su postura respecto al control de las emociones.

Estas dos características (conciencia gradual y posibilidad relativa de control) son las que dan sustento a que la evaluación de la responsabilidad por acciones realizadas en virtud de sesgos implícitos tenga que ser también gradual. El argumento que da para esto es que en las emociones funciona de la misma manera. A continuación, se realizarán algunas observaciones que demuestran que esto no es así –o al menos no siempre– y que los paralelismos entre sesgos y emociones no son tan nítidos, como afirma el autor.

3.2. Estructura de los sesgos y de las emociones: diferencias

En virtud de lo señalado anteriormente, Madva (2017) reconoce una similitud entre la conciencia y el control en los sesgos y en las emociones, y, por ende, conclusiones similares respecto a la responsabilidad que se puede atribuir a quienes realizan acciones emocionales o sesgadas.

Retomemos aquí dos de los rasgos relevantes –que suelen estar presentes–[17] de las emociones identificados por González Lagier: creencia o evaluación y objeto intencional. Los sesgos tienen también estos elementos: hay un objeto o evento (persona o grupo) y una creencia (sobre estas personas o grupos). Sin embargo, la emoción y el sesgo actúan diferente en relación a estos elementos. Veamos este ejemplo:

> X es dueño de un supermercado y está buscando empleados para la sección de reposición de mercaderías. Se presentan a la entrevista A (mujer) y B (varón): X tiene un sesgo implícito, ya que asocia a los varones con actividades de fuerza y a las mujeres con la imposibilidad de realizar tales actividades. X contrata a B.

Es decir:

> Hay un objeto o evento (A y B presentándose al puesto), una creencia sesgada (las mujeres son menos capaces que los varones de realizar actividades que impliquen fuerza) y una acción (B es contratado para el puesto de reposición de mercadería).

Es claro que donde opera el sesgo implícito es en la creencia, es decir, distorsiona la evaluación de un objeto o evento determinado, y luego la persona actúa en función de esa creencia. Lo que puede no ser notado, o al menos no de manera directa, es la posesión de esa creencia sesgada. Es decir, X contrata a B, pero no **porque** los varones son más fuertes que las mujeres, sino porque esa creencia implícita opera en su decisión, aunque él no lo sabe, o no lo sabe de manera directa.

(17) Como ya fue advertido, González Lagier no habla de propiedades necesarias de las emociones, sino de rasgos relevantes o elementos típicos; esto es así porque considera que llamamos emociones a conjuntos de hechos tan disímiles entre sí que no todos comparten todos los rasgos identificados. Utiliza la noción wittgensteiniana de parecidos de familia.

La emoción no actúa necesariamente de esa manera. González Lagier identifica dos papeles distintos de las emociones en la génesis de una acción: contribuyen a formar la intención y limitan causalmente las alternativas de acción. En el primer punto se identifica el papel de la emoción en el entramado de las razones para la acción; en el segundo, se identifica a la emoción como una causa mecánica de la acción. Solo en el primer caso la emoción puede operar sobre la creencia, pero en el segundo caso aparece separada de ella y funciona causalmente. No es lo mismo decir:

> X contrata a B (varón) porque siente desprecio por las mujeres.

Que

> X contrata a B (varón) porque cuando realizó la entrevista a A (mujer) sentía enojo.

En el primer caso, la creencia o evaluación sobre el objeto está afectado por la emoción, mientras que en el segundo caso la emoción afecta el curso de acción, pero no tiene ningún rol en las creencias de X. Madva hace una reconstrucción de las emociones[18] en el segundo sentido (como desconectadas de las creencias), pero las asimila luego a los sesgos en el primer sentido (influyendo en la formación de la creencia). Además, en el primer caso la emoción no funciona **como** un sesgo, sino que es la **causa** de un sesgo, como bien lo sostiene Madva (2017): "Un segundo efecto significativo de los estados de ánimo es que ellos sesgan la cognición, llevando a los individuos selectivamente a prestar atención, ignorar, malinterpretar o percibir erróneamente características de su entorno" (p. 15).[19]

Por otro lado, en su reconstrucción de las emociones a través del ejemplo, Madva no advierte que muchas veces las emociones no solo se utilizan para justificar determinadas acciones y, por consiguiente, atenuar la responsabilidad, sino que también –en ciertas ocasiones– agravan la responsabilidad por esas acciones. No es lo mismo que X diga "no contraté a A porque siento desprecio por las mujeres", que "no contraté a A porque estaba enojado y eso no me permitió prestar atención en la entrevista laboral". En el segundo caso podríamos justificar su acción parcialmente, en el primer caso no; incluso ese desprecio agravaría su responsabilidad. La dificultad de control o limitada conciencia de haber sentido una emoción **no siempre** tiene influencia en la responsabilidad en el sentido de que la mitiga. Esto implica que en la evaluación que hacemos de las acciones emocionales al momento de atribuir responsabilidad, tenemos en cuenta otros factores **además** de la limitación del control y de la conciencia gradual. Recordemos que González Lagier habla de la **intensidad**

(18) Aquí es donde puede verse la diferencia que señalaba anteriormente entre emociones y estados de ánimo en relación a la intencionalidad. Recordemos que Beedie *et al.* (2005) señalaban que las emociones tienen un objeto intencional claro, mientras que eso no sucede en los estados de ánimo. Lo curioso aquí es que Madva toma esa diferencia para reconstruir los estados de ánimo, pero luego los asimila a los sesgos como influyendo en la formación de la creencia, es decir, como asociado a un objeto intencional, desdibujándose allí la distinción entre emociones y estados de ánimo.

(19) La traducción me pertenece.

de la emoción y la **justificación de la creencia.** Lo primero tiene que ver con la conciencia y el control (intensidad), en tanto que lo segundo es externo y tiene que ver con que la creencia esté justificada (por su razonabilidad, proporcionalidad, etc.). Esto último no es tenido en cuenta por Madva al momento de analizar la responsabilidad.

Por último, Madva, al realizar este paralelismo entre emociones y sesgos, resalta los casos en los que la persona que tiene una emoción no se da cuenta de que la tiene, y asimila eso a la portación de un sesgo. Sin embargo, hay muchos otros casos en los que las personas realizan acciones emocionales aun sabiendo que tienen esa emoción, lo que permitiría –en algún sentido– controlar más esas acciones. Por lo tanto, los sesgos implícitos solo podrían asemejarse a los casos en los que las emociones no son percibidas.

Entonces, el autor parte de una errónea comprensión de las emociones –o al menos parcial–, lo que hace que los paralelos que traza entre ambos fenómenos sean también erróneos o parciales. La limitada conciencia y el limitado control no son condiciones necesarias para la existencia de las emociones. Además, no hay una relación directa entre conciencia, control y responsabilidad como parece afirmar Madva (2017), por lo menos no para las teorías que no son mecanicistas (las que él explícitamente rechaza en su artículo). Es necesario complejizar el análisis y determinar qué otros factores o condiciones entran en juego al momento de analizar la responsabilidad por emociones. La misma consideración vale para el caso de los sesgos.

Las emociones son fenómenos complejos que actúan de modos diversos sobre las creencias y acciones de los individuos. Tienen propiedades que no están presentes en los sesgos implícitos, como las sensaciones físicas, y algunas que, aunque pueden parecer similares, no necesariamente lo son. No es posible trazar un paralelo tan claro y sacar conclusiones respecto de los sesgos implícitos debido a que, por un lado, no comparten muchas de las características relevantes y que, por otro, las discusiones dentro de las teorías de las emociones lejos están de haber terminado. Sin embargo, esto no significa que los estudios sobre los sesgos no tengan cosas que aprender de las teorías de las emociones. El camino, no obstante, no es tan simple como la asimilación de los dos fenómenos.

4. A modo de conclusión: ¿qué pueden aportar las teorías de las emociones a los estudios sobre sesgos implícitos?

Existen, como vimos, algunas confusiones conceptuales en las pretensiones de trazar paralelos entre emociones y sesgos. Esto no implica, sin embargo, un abandono de las teorías de las emociones para abordar los sesgos. Es claro que es necesario manejarse con cautela en el terreno de los sesgos porque los estudios se encuentran mucho menos avanzados, e incluso hay dudas sobre si es posible demostrar su existencia.[20]

(20) Para ampliar sobre este punto, véase Mitchell y Tetlock (2006).

Como se expresó a lo largo de este trabajo, lo que se entiende por emoción tiene directa relación con la responsabilidad que se les puede atribuir a las personas por la realización de una acción emocional. Esto vale también para los sesgos.

La forma en que suele analizarse la responsabilidad implica un vínculo entre la acción concreta realizada y los estándares externos a los que esa acción debe adecuarse. Es decir, la respuesta sobre la voluntariedad, conciencia o controlabilidad de la acción no es suficiente para fundar el juicio sobre la responsabilidad, sino que además se evalúa la acción de acuerdo con parámetros sobre lo que sería correcto hacer en esa situación determinada. En otras palabras, se evalúa si la acción realizada es lo que cualquier persona **razonable** habría hecho si se encontrara en la misma situación, o si se basa, además, en creencias razonables o justificadas. Estos estándares incluyen cualidades generales sobre los roles que ocupan los sujetos o las tareas que están cumpliendo y valoraciones o evaluaciones sobre cómo se debe actuar y cuáles deben ser las razones para nuestras acciones. No se aplica el mismo parámetro, por ejemplo, a cualquier ciudadano/a que a un juez o jueza, que tienen la obligación legal de regirse por determinadas reglas de actuación. Las teorías mixtas de las emociones atienden estas cuestiones, y puede resultar útil esta distinción también para preguntarse sobre los efectos de los sesgos implícitos en la responsabilidad.

No hay dudas de que las teorías de las emociones pueden orientar de muchas maneras las discusiones sobre los sesgos implícitos, pero esto no puede quedar reducido a esfuerzos por asimilar ambos fenómenos. Por el contrario, las similitudes y diferencias entre ellos pueden servir para repensar la relación entre las propiedades internas de cada uno con las consecuencias prácticas de las mismas. En este sentido, que atribuyamos responsabilidad parcial o gradual por acciones emocionales y acciones sesgadas puede deberse a razones diversas, y que tengan la misma consecuencia práctica no necesariamente tiene que ver con una similitud de sus propiedades.

Una teoría analítica más completa de los sesgos implícitos debe empezar por aportar claridad conceptual respecto a qué son, cuáles son sus propiedades y cómo dar pruebas de su existencia, para luego pasar a pensar qué consecuencias prácticas tienen en las acciones y decisiones de las personas y, por ende, en la responsabilidad. No deben obviarse las complejidades del abordaje del fenómeno en los distintos niveles.

5. Referencias bibliográficas

Ahmed, S. (2015). *La política cultural de las emociones.* C. Olivares Mansuy (Trad.). Universidad Nacional Autónoma de México/Centro de investigaciones/Estudios de Género.

Beedie, C. J.; Terry, P. C. y Lane, A. M. (2005). Distinctions between emotions and moods. *Cognition and Emotion, 19*(6), pp. 847-878.

Canale, D.; Ciuni, R.; Frigerio, A. y Tuzet, G. (Eds.). (2021). *Critical Thinking: an introduction.* EGEA Spa, Bocconi/University Press.

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. *Revista Derecho Penal y Criminología, 38*(102), pp. 141-164. dx.doi.org/10.18601/01210483.v37n102.08

Descartes, R. (1997). *Las pasiones del alma.* J. A. Martínez Martínez y P. Andrade Boué (Trads.). Tecnos.

Duff, R. A. (2015). Criminal responsibility and the emotions: if fear and anger can exculpate, why not compassion? *Inquiry: An Interdisciplinary Journal of Philosophy, 58*(2), pp. 189-220. doi.org/10.1080/0020174X.2015.986855

Frankish, K. (2016). Playing double. Implicit bias, dual levels and self-control. En M. Brownstein y J. Saul (Eds.), *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology.* Oxford Scholarship Online.

González Lagier, D. (2009). Los presupuestos de la responsabilidad por nuestras emociones, *Doxa, Cuadernos de Filosofía del derecho, 34*, pp. 439-458.

----------. (2010). *Emociones, responsabilidad y derecho.* Marcial Pons.

Holroyd, J.; Scaife, R. y Stafford, T. (2017). What is implicit bias? *Philosophy Compass, 12:e12437.* doi.org/10.1111/phc3.12437

Stafford, T.; Holroyd, J. y Scaife, R. (2018). Confronting bias in judging: a framework for addressing psycologhical biases in decision making. doi.org/10.31234/osf.io/nzskm

Kahan, D. M. y Nussbaum, M. C. (1996). Two Conceptions of Emotion in Criminal Law. *Columbia Law Review, 96*(2), pp. 270–374.

López, H. (2017). Prólogo. En S. Ahmed, La política cultural de las emociones. C. Olivares Mansuy (Trad.). Universidad Nacional Autónoma de México/Centro de investigaciones/Estudios de Género.

Madva, A. (2017). Implicit bias, moods and moral responsibility. *Pacific Philosophical Quarterly.* doi.org/10.1111/papq.12212

Manrique, L. (2018). Emociones, acción y excusas. *Eunomía. Revista en Cultura de la Legalidad, 14*, pp. 71-86. doi.org/10.20318/eunomia.2018.4156

Mitchell, G. y Tetlock, P. E. (2006). Antidiscrimination Law and the Perils of Mindreading. *Ohio State Law Journal, 67*, pp. 1023-1122.

Nussbaum, M. C. (2008). *Paisajes del pensamiento: La inteligencia de las emociones. Magnum: Vol. 2.* Paidós.

Vincent, N. A. y Nadelhoffer, T. (2013). *Neuroscience and Legal Responsibility.* Oxford University Press.

Uniacke, S. (2007). Emotional excuses. *Law and Philosophy, 26,* pp. 95-117. doi.org/10.1007/s10982-006-0003-y