

Cognición Moral

María C. Castro
Santiago Roa
Santiago Amaya
Universidad de los Andes

Los últimos veinte años han sido testigos del surgimiento de la psicología moral como un campo común de investigación entre la filosofía y las ciencias cognitivas. En términos generales la psicología moral investiga cómo piensan y cómo se comportan las personas en contextos moralmente relevantes (Lapsely, 1996; Doris y Stich 2014; Doris et al., 2010). Es decir, sus preguntas tienen que ver con las capacidades humanas para construir una realidad moral, comprenderla y transformarla.

Ciertamente, este tipo de preguntas no son nuevas; ellas han sido provincia tanto de la ética normativa como de la psicología social. La psicología moral contemporánea, sin embargo, aborda estas preguntas combinando metodologías diversas: desde el uso de técnicas de análisis conceptual y experimentos mentales, hasta estudios de resonancia magnética funcional. Como veremos, detrás de esta promiscuidad metodológica, hay una nueva manera de trazar las líneas que usualmente han servido para dividir lo normativo de lo descriptivo.

Los psicólogos morales se han ocupado de una multiplicidad de temas: la motivación, los sentimientos morales, la evolución de la moralidad, el desacuerdo, el carácter moral, entre otros. Es imposible en este breve capítulo hacer justicia a este amplio rango (Un panorama más amplio puede verse en Sinnott-Armstrong, 2008a; 2008b; 2008c; 2014 y Doris et al., 2010). En vez de intentar hacerlo, aquí nos concentraremos en un tema central, el de la *psicología del juicio moral*, intentando ilustrar la variedad de perspectivas y la riqueza de la reflexión y experimentación

en este campo. En concreto, nos concentraremos en lo que aquí llamaremos respectivamente “juicios de estatus moral” y “juicios de responsabilidad”.

Vale la pena comenzar con un par de aclaraciones y de precisiones de corte metodológico. Ellas servirán para contextualizar la discusión que sigue y para ubicar a la psicología moral en el horizonte de algunas disputas propias de la filosofía de las ciencias cognitivas.

1. Juicio moral: normativo/descriptivo

La capacidad general de hacer juicios morales, es decir, de hacer evaluaciones morales, es central para el funcionamiento de las personas y su realidad moral. En algunos casos, los juicios morales son determinantes de las decisiones y de la conducta. En otros, sirven como expresiones de los estándares comunes alrededor de los cuales se articula la vida moral de una comunidad. De manera general, para asegurar la cooperación y la pro-socialidad, los individuos y comunidades deben poder evaluar ciertos comportamientos y actitudes, ya sea para promoverlos o para castigarlos.

Los juicios morales pueden ocuparse de diferentes objetos y atribuirles a éstos diferentes tipos de propiedades. De ahí que no haya algo como *el* juicio moral; más bien, hay diferentes tipos de juicios morales (Cushman, 2008; Malle, Guglielmo y Monroe, 2014). En este capítulo nos concentraremos en dos de ellos. En primer lugar, están los juicios sobre el *estatus moral* de un determinado comportamiento. “Robar está mal”, “ayudar está bien”, “mentir es indebido”, son ejemplos de este tipo de juicios. En segundo lugar, están los juicios sobre la *responsabilidad moral* de un agente con respecto a una determinada conducta. Estos juicios suelen tener la forma “S es responsable de ϕ ” o incluso “S merece sanción por haber hecho ϕ ”.

Por su relevancia en la vida moral de individuos y grupos es importante comprender los determinantes del juicio moral. Si bien la cognición moral no se limita a la formulación de juicios, ellos son una ventana hacia ésta. En la psicología del juicio confluyen diferentes tipos de razonamiento moral, ella está determinada por procesos tanto cognitivos como afectivos y varias de las categorías y conceptos que estructuran dichos juicios atraviesan la cognición moral en general.

Ahora bien, tradicionalmente, la ética normativa y la filosofía de la acción se han ocupado de los juicios morales. Su tarea ha sido mayoritariamente *prescriptiva* (Driver, 2007; Shafer-Landau, 2014). Es decir, han buscado establecer qué hace que estas atribuciones (de estatus o de responsabilidad) sean correctas o verdaderas y cómo pueden ser ellas justificadas. En menor medida su interés ha sido cómo las personas hacen *de facto* estas atribuciones o cómo las justifican, independientemente de qué tan correctas sean éstas. Desde luego, esto no significa que dichas teorías no hagan de rutina presuposiciones (algunas incluso controversiales) sobre la psicología de agentes morales como los seres humanos (ver al respecto la discusión en Doris y Stich, 2014).

Algunas doctrinas filosóficas parecerían sugerir que este proyecto normativo puede (o debe) proceder independientemente del tipo de investigación descriptiva y experimental propia de la psicología moral. Un “debe”, según David Hume (1739/2005; 1751/2006), no puede *derivarse* de un “es”. Es una falacia, según G.E. Moore (1903/1993), intentar *analizar* una categoría o propiedad moral en términos de una categoría o propiedad natural (más aún en términos psicológicos). Obviamente, estas doctrinas pueden ser y han sido cuestionadas (dos discusiones clásicas son Anscombe (1958) y Searle (1964)). Pero, más allá de estas críticas, la realidad es que una gran mayoría de psicólogos morales consideran que hay una estrecha relación entre la reflexión normativa y descriptiva con respecto al juicio

moral. Después de todo, derivación y análisis son solo dos de muchas posibles formas en que la ética normativa y la psicología moral podrían relacionarse.

Consideremos los juicios de estatus moral. Siguiendo a John Rawls (1955), quien a su vez toma la idea de Nelson Goodman (1955), varios filósofos piensan que la adopción de principios morales debe estar guiada por un ideal de *equilibrio reflexivo* (una discusión del equilibrio reflexivo puede encontrarse en Daniels (1979) y en Vargas (2007)). Puesto brevemente, la idea es que dichos principios son aceptables en la medida en que permitan explicar y racionalizar *intuiciones* morales básicas, es decir, juicios morales particulares comúnmente aceptados. Estas intuiciones, por así decirlo, son los datos de los que una teoría normativa moral debe dar cuenta.

Sin embargo, como veremos más adelante, el estudio de la psicología del juicio moral muestra que algunas de estas intuiciones deben aproximarse con escepticismo. Algunas de ellas son más el producto de los métodos de evocación usados que de convicciones personales (Knobe y Doris, 2010). Otras son evidencia de la existencia de “efectos de marco” en el juicio moral (Sinnott-Armstrong, 2008b). Finalmente, muchas de ellas parecen ser meras racionalizaciones post-hoc (Nichols, 2014; Schwitzgebel y Ellis, 2017). Es decir, si las intuiciones morales son datos para una teoría normativa, la evidencia de la psicología moral tiene el potencial para desacreditar algunas intuiciones en tanto datos guía para el establecimiento de principios normativos generales.

Algo similar sucede en el caso de los juicios de responsabilidad. Tradicionalmente, al menos desde Kant (1781/2011), muchos filósofos han argumentado que la responsabilidad moral depende de la cuestión metafísica sobre la libertad de la voluntad. Solo hay responsabilidad, por ejemplo, si el determinismo causal *no* es cierto. Contrario a esta tradición, varios compatibilistas contemporáneos han insistido en que la existencia de la responsabilidad moral no

debe depender de cuestiones metafísicas como ésta. Tomando como inspiración a Peter Strawson (1962/2013), han argumentado que la existencia de la responsabilidad moral puede ser tomada como un hecho constituido por nuestras prácticas comunes de atribución de responsabilidad (Algunos ejemplos pueden encontrarse en Watson (2013), Vargas (2013a) y Shoemaker (en prensa)). Una teoría normativa de la responsabilidad, en parte, debe poder explicar estas prácticas.

Hay aquí, entonces, otra manera en que la psicología moral puede contribuir al desarrollo de una teoría filosófica normativa. Algunos psicólogos morales han investigado hasta qué punto el concepto cotidiano de responsabilidad moral presupone o no la falsedad del determinismo causal—de hacerlo, algunas posiciones compatibilistas estarían en problemas (Nahmias, Morris, Nadelhoffer y Turner, 2005; Woolfolk, Doris y Darley, 2006; Nichols y Knobe, 2007; Knobe y Doris, 2010). De manera más general, el estudio de la psicología de los juicios de responsabilidad es un primer paso para comprender las prácticas comunes de atribución de responsabilidad que, desde esta perspectiva, la teoría normativa debe explicar.

Obviamente, no solo la reflexión filosófica tradicional puede enriquecerse del trabajo desarrollado por psicólogos morales. También la influencia positiva corre en la dirección contraria. Simplemente, por ser éste un volumen sobre la *filosofía* de las ciencias cognitivas hemos enfatizado esta dirección. En todo caso, como veremos más adelante, las voces centrales en los debates actuales de la psicología moral incluyen a profesionales en filosofía, psicología social y muchas otras de las ciencias cognitivas. Sea éste un recordatorio de que el término “psicología” en el rótulo “psicología moral” marca más una serie de temáticas de estudio que una unidad en la repartición de la torta académica burocrática de los siglos XX y XXI.

2. Juicios de estatus

Una de las áreas centrales de estudio en la psicología del juicio moral ha sido lo que aquí llamamos “juicios de estatus”. Estos son juicios en los que una acción (o una omisión) es evaluada positiva o negativamente a través de la atribución de una determinada propiedad moral. Es difícil decir qué hace que una propiedad sea moral o no. De hecho, diferentes teorías y perspectivas normativas varían con respecto a qué propiedades se consideran morales y cuáles de estas tienen prelación en el juicio (Para la discusión de algunas posibilidades ver Driver (2007) y Shafer-Landau (2014)).

En general, la psicología moral contemporánea se ha enfocado en atribuciones de bueno/malo o correcto/incorrecto. Es posible que esto se deba a que éstas son las categorías morales más generales y que, por tanto, los resultados obtenidos con respecto a ellas sean generalizables a juicios que involucran otras categorías, por ejemplo, lo permisible vs. lo obligatorio (Bjorklund, 2003; Koenigs, Kruepke, Zeier y Newman 2012; O’Hara, Sinnott-Armstrong y Sinnott-Armstrong, 2010). Pero ésta es una hipótesis empírica con respecto a la cual no hay evidencia concluyente (Barbosa y Jiménez, 2017). En todo caso, es imposible en el estudio del juicio moral no presuponer una cierta comprensión pre-teorética de la moralidad, lo cual se evidencia en lo que parecerían ser sesgos en algunas de las propuestas que discutiremos a continuación. Como veremos, uno de los retos más visibles a este respecto tiene que ver con el diseño de las herramientas para el estudio experimental de los mecanismos detrás del juicio moral.

Para efectos de nuestra exposición, comenzaremos discutiendo lo que podrían denominarse aproximaciones *racionalistas o cognitivistas* al juicio moral. Seguidas de estas, introduciremos algunas teorías de corte más *intuicionista o sentimentalista*. Vale la pena aclarar que nuestro énfasis en este capítulo está

primariamente en aproximaciones experimentalmente motivadas, lo cual significa que dejaremos de lado algunas de las teorías tradicionales que de otro modo han hecho carrera en la filosofía moral. Incluso, a este respecto, nuestra discusión estará limitada alguna de las aproximaciones que pueden considerarse representativas.

2.1. Racionalismo

Los orígenes de la tradición racionalista o cognitivista pueden rastrearse, al menos, a las ideas del Immanuel Kant (1785/2002), para quien el estatus moral de una acción era una función de su posible derivación de mandatos de la facultad de la razón práctica. Aunque en la actualidad hablar de una “facultad” de la razón ha caído en desuso, hay una variedad de posiciones que pueden considerarse racionalistas. La idea general que las cobija no es que el juicio moral sea el producto de una llamada facultad (aunque más abajo mencionamos una versión de esta idea), sino que es el proceso de la aplicación de reglas o principios generales para evaluar las acciones.

Lawrence Kohlberg ocupa un lugar privilegiado dentro de la tradición experimental racionalista, tanto por los métodos que desarrolló como por sus planteamientos sobre desarrollo moral (Kohlberg, 1984/1992; Kohlberg y Hersh, 1977; Colby et al., 1983). Siguiendo las ideas de Piaget (1932/1965), Kohlberg se propuso estudiar el desarrollo moral en el niño caracterizando los cambios en su juicio moral. Su hipótesis era que dichos cambios surgían como consecuencia de la adquisición de formas de razonamiento cada vez más sofisticadas.

Para estudiar esta hipótesis, Kohlberg diseñó un estudio compuesto de dilemas morales hipotéticos (inspirados en ejemplos de la *Ética a Nicómaco* de Aristóteles), seguidos de una serie de cuestionarios con los que se buscaba evocar en los participantes (niños de alrededor de los 10 años) juicios morales y tipos de

razonamiento asociados a ellos. Los dilemas describen situaciones en las que existen potencialmente violaciones de deberes, conflictos entre deberes y derechos o entre distintos tipos de deberes. Como veremos en el transcurso de esta discusión, este tipo de dilemas presentados en forma de viñetas han sido una herramienta central en la investigación sobre el juicio moral.

Basado en los resultados obtenidos, Kohlberg propuso dividir el desarrollo del juicio moral en diferentes estadios (6 en total). Haciendo eco de la distinción kantiana entre juicios heterónomos y autónomos, cada uno de ellos se caracteriza por la presencia de razonamientos cada vez más impersonales en la formulación del juicio. Estos estadios van desde lo pre-convencional (“hacer ϕ está mal porque podrían castigarme por ello”), pasando por lo convencional (“hacer ϕ está mal porque está prohibido”), hasta lo posconvencional. Este último se caracteriza por la presencia de principios de justicia enmarcados en términos de prescripciones universales: “hacer ϕ está mal porque es una violación de x o y derechos” (Kohlberg y Hersh, 1977; Kohlberg, 1984).

Un presupuesto central en el planteamiento de Kohlberg, evidente en el tipo de dilemas utilizados por él, es que el desarrollo del juicio moral puede trazarse en términos de la aparición de una normatividad definida sobre derechos y deberes. Sin embargo, como hemos visto, los juicios de estatus moral incluyen atribuciones de una variedad de propiedades y no es claro que consideraciones *deontológicas* como éstas, que desde luego juegan un papel en el juicio moral, tengan una cierta preeminencia al respecto. Precisamente, algunas de las objeciones centrales que ha recibido la teoría de Kohlberg han surgido de una insatisfacción a este respecto. Como estos críticos lo han enfatizado, el dominio de la moralidad y, por tanto, del juicio moral se extiende a otras formas de valoración (Dos críticas clásicas pueden encontrarse en Gilligan (1982; 1995) y Eisenberg (1979; 1986)).

Si bien la influencia de Kohlberg en la psicología del desarrollo es innegable, su influencia en la psicología moral contemporánea es comparativamente menos directa (para discusión ver Lapsley (2005) y Lapsley y Narvaez (2005)). No por ello es menos importante. Además del uso de dilemas hipotéticos como herramienta central en el estudio del juicio moral, varias de las ideas centrales a sus planteamientos han re-aparecido con fuerza en la discusión actual entre psicólogos morales. A continuación, discutimos brevemente una de ellas: la distinción entre normas convencionales y normas morales.

2.2. Convencional vs. Moral

Aunque Kohlberg no plantea propiamente un contraste entre tipos de normas sino entre etapas de desarrollo, su distinción entre el nivel convencional y el posconvencional anticipa el interés contemporáneo en la distinción entre normas convencionales y morales (Huebner, Lee y Hauser, 2010; Levy, 2005; O'Neill, 2017). Este contraste ha sido una de las herramientas conceptuales que la psicología moral contemporánea ha usado para definir el dominio de la cognición moral. Hay distintas propuestas acerca de cómo podemos entender esta diferencia. Aun así, una manera general de entenderla es que las normas convencionales se encargan de prohibiciones arbitrarias que permiten la organización y la coordinación social, mientras que las normas morales se encargan principalmente de prohibiciones acerca del daño (Kelly, Stich, Haley, Eng y Fessler, 2007). Desde luego, esta formulación es de por sí ya contenciosa: el campo de lo moral es presumiblemente más amplio e incluye también dominios como el de la lealtad, la autoridad y la pureza espiritual (Haidt, 2007).

Elliot Turiel es el pionero del intento contemporáneo por entender, desde un punto de vista cognitivo, la diferencia entre estos tipos de normas (Blair, 1993;

Nucci, 1986; Smetana y Braeges, 1990; Smetana, Schlagman y Walsh, 1993). Según él, lo que distingue las normas morales de las convencionales es su contenido y la forma en que éstas son justificadas. Las normas morales son acerca del bienestar, la justicia y los derechos (Turiel, Killen y Helwig, 1987); su justificación tiene que ver con la seriedad de sus violaciones, su independencia de la autoridad y el sufrimiento de otros (Turiel, 1983). Las normas sociales, por otro lado, codifican uniformidades en las interacciones sociales y están justificadas por el funcionamiento del sistema social correspondiente (Nucci y Turiel, 1978).

Hoy en día, la distinción entre estos tipos de normas se estudia con tareas que parten del diseño que Turiel propuso originalmente (Ver por ejemplo Kelly et al. (2007), Mendez, Anderson y Shapira (2005) y Nichols (2002)). En ellas los participantes reciben una descripción de distintos escenarios en los que hay o bien una transgresión convencional (violaciones a normas escolares, de etiqueta o familiares) o bien una transgresión moral (halar el pelo de alguien más, robar o empujar a alguien). Con relación a éstas, los participantes deben responder a una serie de preguntas que se consideran indicativas de cómo trazan ellos la distinción entre trasgresiones morales y convencionales. Por ejemplo ¿qué tan severamente deben ser castigados los transgresores? ¿si la autoridad lo permite, está bien cometer la transgresión?

Una hipótesis interesante es que a la distinción entre normas morales y convencionales subyace una diferencia que va más allá de su contenido y posibles justificaciones (Nichols 2002; 2004; Prinz, 2006). Shaun Nichols (2002; 2004), por ejemplo, ha propuesto que los juicios convencionales son el producto de la aplicación de principios propios de una teoría normativa acerca de lo que socialmente ha sido acordado. Los juicios morales, por el contrario, no solo son aplicaciones de una teoría normativa sino el resultado del funcionamiento de un mecanismo afectivo que registra el sufrimiento de otras personas y que se encarga

de generar reacciones apropiadas a transgresiones percibidas. En este orden de ideas, la distinción entre lo moral y lo convencional sería evidencia de una distinción entre dos tipos de arquitecturas mentales para el procesamiento de normas.

Siguiendo con la tradición de Kohlberg, esta distinción se ha estudiado mayoritariamente en poblaciones infantiles como un indicador del desarrollo de sus estructuras cognitivas (Blair, 1993; 1997; Dunn y Munn, 1987; Nucci, 1986; Nucci y Turiel, 1978; Smetana, 1989; Smetana y Braeges, 1990). Pero las preguntas también se han adecuado para casos más propios de la adultez y los resultados han demostrado que los adultos también distinguen las trasgresiones morales de las convencionales siguiendo los criterios anteriormente mencionados (Hauser, Tonnaer y Cima, 2010; Huebner et al., 2010).

Más interesante aún, esta capacidad se ha estudiado en poblaciones con distintos desordenes psiquiátricos, en particular, en pacientes con distintas formas de psicopatía. Algunos estudios han reportado un déficit en la capacidad de éstas personas para hacer la distinción moral/convencional o, al menos, para verse motivadas por ella (Blair y Cipolotti, 2000; Kinga et al., 2005). En general, en comparación con los resultados observados en niños y adultos neurotípicos, los psicópatas parecen acudir con más frecuencia a justificaciones de tipo eminentemente convencional, aún para transgresiones que implican un daño físico a un tercero, lo que sugiere que la distinción moral/convencional tiene que ver con maneras distintas de procesar normas de comportamiento (Blair 1995; 1997).

Si bien es cierto que la distinción entre lo moral y lo convencional ha jugado un papel en el estudio del juicio moral, ella también ha sido cuestionada. De modo general, sus detractores han intentado demostrar que las características con las que se ha pretendido delimitar la distinción no son exclusivas de uno u otro ámbito (Nisan, 1987; Nucci y Turiel, 1993; Nichols, 2002; 2004). Por un lado, algunos han

demostrado que existen algunas trasgresiones que suelen considerarse como morales, el incesto por ejemplo, aun cuando no se justifican acudiendo al bienestar de otro sino al disgusto que generan (Haidt, 2001). También hay evidencia de que el estatus socio-económico de las personas afecta cómo se traza en concreto la distinción entre lo moral y lo convencional: el incesto tiende a ser visto más como una trasgresión meramente social (es decir, no moral) en estratos altos en comparación con estratos bajos (Haidt, Koller y Dias, 1993). Finalmente, algunas trasgresiones se consideran morales dependiendo del momento histórico en el que ocurrieron. Es decir, al evaluarse estas acciones como transgresiones morales influye no sólo su contenido o el tipo de justificación asociada a ellas, sino también el contexto normativo relevante (Kelly et al., 2007).

2.3. La analogía lingüística

Otra gran fuente de influencia para las teorías racionalistas sobre el juicio moral puede encontrarse en las ideas lingüísticas de Noam Chomsky (ver por ejemplo Chomsky (1957; 1986)). Siguiendo una sugerencia del filósofo John Rawls (1971), varios psicólogos morales han argumentado que existe una estrecha analogía entre el conocimiento lingüístico (especialmente sintáctico), tal como Chomsky lo entiende, y el conocimiento moral. De acuerdo con esta posición, así como los juicios sobre la gramaticalidad de ciertas construcciones entre hablantes nativos pueden explicarse en términos del conocimiento implícito de reglas sintácticas y habilidades computacionales, así también las intuiciones comunes sobre la moralidad de ciertas acciones pueden verse como la manifestación de dichas habilidades y del conocimiento *tácito* de ciertos principios de justicia y bienestar (Harman, 1999; Dwyer, 1999; 2006; Mikhail, 2000).

Existen, sin embargo, varias maneras de interpretar la analogía entre conocimiento lingüístico y conocimiento moral (Hauser, Young y Cushman, 2008) y no es claro que todos sus defensores tengan una versión unificada en mente. Una versión fuerte, por un lado, supondría que en el campo de la moral existen principios análogos a las reglas universales de la gramática chomskiana, que la diversidad en el juicio moral está limitada en el mismo sentido en que la variabilidad sintáctica parece estarlo y que el procesamiento de principios morales ocurre en lugares dedicados de la corteza cerebral (una facultad de la moralidad, por analogía a las llamadas áreas del lenguaje: Broca y Wernicke). Sobra decir que en el momento la evidencia no es suficiente para defender esta versión fuerte de la analogía. Obviamente también diferentes aspectos de esta versión fuerte pueden disociarse unos de otros.

En contraste, versiones débiles asumen la analogía como un recurso heurístico para pensar la cognición moral desde una perspectiva racionalista, en particular, como un recurso para pensar la cognición moral como un proceso computacional basado en reglas. En este sentido, Cushman y sus colegas (Mikhail (2002), Cushman, Young y Hauser (2006)) han diseñado una serie de experimentos con dilemas morales basados en el famoso caso del tranvía (Foot, 1967; Thomson, 1976; 1985). Su objetivo ha sido establecer cuáles son los principios generales de enjuiciamiento detrás de las intuiciones morales comunes que este dilema suscita y que los participantes por sí mismos no necesariamente pueden articular. Interesante a este respecto, los resultados sugieren que entre estos principios se encuentran algunos ampliamente debatidos en la tradición de la ética normativa, por ejemplo, el principio del doble efecto (Para una discusión de este principio ver Mikhail (2002), McIntyre (2014)).

2.4. Afecto e intuición

Por comparación con el racionalismo, otra manera de aproximarse a los juicios de estatus corresponde a lo que en términos filosóficos clásicos podría llamarse la tradición sentimentalista. De acuerdo con ésta, los juicios sobre el estatus moral de nuestras acciones dependen, no tanto de mecanismos generales de razonamiento y del conocimiento tácito o explícito de principios morales, sino de una cierta sensibilidad emocional. Esta propuesta se puede rastrear hasta las ideas del filósofo David Hume (1739/2005; 1751/2006), según él cual los juicios morales son expresiones de *sentimientos* de aprobación o desaprobación hacia aquello que juzgamos.

Las teorías sentimentalistas contemporáneas vienen en varias versiones. Algunas de ellas tienen una clara raigambre filosófica humeana (Algunos ejemplos pueden encontrarse en Nichols (2002), D'Arms (2005), Prinz (2006; 2008)). Otras han estado motivadas más por propuestas generales en las ciencias cognitivas o por algunos resultados experimentales particulares (Greene 2008; 2009; 2013; Greene y Haidt 2002; Prinz, 2006; 2008; Haidt 2001; 2007). Aquí nos ocuparemos sólo de algunas de ellas.

Uno de los principales proponentes del sentimentalismo en la psicología moral reciente ha sido Shaun Nichols. Como ya vimos, en el modelo de Nichols las reacciones emocionales juegan un rol importante en la distinción entre normas morales y convencionales. De modo más general, según su modelo, las intuiciones morales son el resultado de la operación de dos tipos de elementos. Está, por un lado, una teoría normativa implementada en un set de información que especifica qué tipo de acciones deben considerarse erróneas. Por otro lado, hay un mecanismo afectivo que se activa con la trasgresión de dichas reglas y que responde primariamente a instancias de sufrimiento ajeno.

Además de consideraciones filosóficas sobre la naturaleza de la motivación moral en general, la teoría sentimentalista de Nichols responde también a la necesidad de explicar déficits documentados que algunas poblaciones demuestran en su habilidad para formular juicios morales. Al no haber asimilado todavía una teoría normativa, por ejemplo, los niños menores de un año tienen dificultades haciendo la distinción entre lo moral y lo convencional (Smetana y Braeges, 1990). De igual forma, como mencionamos anteriormente, si bien los psicópatas en general entienden de qué se trata la distinción, esta comprensión tiende a no ser relevante desde el punto de vista de sus decisiones y justificaciones. Esto sugiere, y así parece confirmarlo la evidencia, que ellos tienen la información correspondiente a la teoría normativa, pero un déficit en su reacción afectiva al sufrimiento (Damasio, 1995; Bechara, Damasio y Damasio, 2000; Nichols, 2002; Roskies, 2003).

Otra teoría sentimentalista de corte bastante diferente puede encontrarse en las ideas de Jonathan Haidt (2001). Su modelo del *intuicionismo social* se contrasta con el de Nichols en al menos dos aspectos. Por un lado, Haidt difiere con respecto a la idea de que los juicios morales están basados en una teoría normativa que especifica reglas de comportamiento. Por otro lado, a diferencia de una gran parte de la psicología moral en la que la moralidad suele restringirse al ámbito del daño, Haidt considera que el ámbito moral es más amplio y que incluye dimensiones más amplias como la repugnancia y la lealtad.

Un impulso general al modelo de Haidt proviene de su trabajo con relación al fenómeno conocido como la “perplejidad moral” (*moral dumbfounding*). En una serie de estudios (Haidt, Björklund y Murphy, 2000), Haidt presentó a sus participantes una serie de viñetas que describen trasgresiones en las que no se comete un daño aparente (canibalismo o incesto), pero que, aun así, suelen ser consideradas por éstos como moralmente incorrectas. El resultado general es que, a

pesar de estar seguros de la inmoralidad de las acciones descritas, los participantes suelen ofrecer justificaciones que no son acordes con el escenario descrito (porque citan posibles daños cuando de hecho no los hay) o suelen terminar reconociendo que, a pesar de estar convencidos de su inmoralidad, no saben por qué condenaron la acción.

Según Haidt, lo que estos casos (y otros similares muestran) es que el juicio moral normalmente está precedido por intuiciones, no tanto por un razonamiento basado en reglas y principios. Dichas intuiciones son evaluaciones con una valencia afectiva. Ellas son rápidas y automáticas y, al ser verbalizadas, se convierten en juicios morales (Haidt, 2001). El razonamiento moral, por contraste, solo opera en ciertos contextos sociales en los que, por una u otra razón, se exige la verbalización y justificación de dichas intuiciones. Incluso allí solo tiene una función justificatoria *post hoc*. Es decir, el razonamiento moral y los principios sobre los que éste versa no son, en la propuesta de Haidt, parte propia de la etiología del juicio, sino capacidades y expedientes útiles para la deliberación pública sobre nuestras opiniones morales y para convencer a otros sobre su verdad.

2.5. Procesos Duales

La hipótesis de los procesos duales ha sido una de las perspectivas sobre arquitectura cognitiva de más figuración en las ciencias cognitivas contemporáneas (Evans, 1989; Evans y Stanovich, 2013; Kahneman, 2011; Wason y Evans, 1975). Según esta hipótesis para cada tipo de tarea cognitiva general pueden encontrarse al menos dos procesos cualitativamente distintos encargados del procesamiento de información, los cuales frecuentemente compiten entre sí. Uno de ellos (el procesamiento tipo 1) es rápido y automático, mientras que el otro (el

procesamiento tipo 2) es lento y deliberativo. Esta hipótesis acerca de nuestra arquitectura mental ha sido utilizada para conceptualizar resultados en áreas tan disímiles como el aprendizaje, la categorización y la cognición social. En tanto proceso cognitivo general, la hipótesis también ha sido aplicada en el dominio del juicio moral.

Joshua Greene (2008) ha desarrollado una versión de esta idea con el objetivo de conceptualizar varios resultados observados a propósito de estudios con el dilema del tranvía. En general, dichos resultados parecen sugerir la existencia de dos formas de pensamiento moral que él caracteriza como “utilitarista” y “deontológico,” en alusión a las corrientes en filosofía moral que dichas formas de pensamiento evocan. Puesto brevemente, en una de ellas las acciones son evaluadas como buenas o malas en función de sus consecuencias; en la otra la evaluación depende de prohibiciones categóricas generales. De acuerdo con Greene a estas formas de pensamiento subyacen dos tipos de procesamiento cognitivo. El primer tipo de juicio está guiado por procesos tipo 2, mientras que el segundo está guiado por procesos tipo 1 (Cushman, Young y Greene, 2010).

Varias formas de evidencia han sido presentadas en favor de la idea de que estas dos formas de pensamiento, en efecto, encarnan diferentes tipos de procesos cognitivos. En primer lugar, cada una de estas formas de pensamiento parece tener una estructura temporal diferente. Los juicios que reflejan un pensamiento no consecuencialista suelen ser considerablemente más rápidos que los que sí lo son (Greene, Sommerville, Nystrom, Darley y Cohen, 2001). Segundo, en concordancia con la hipótesis general de los procesos duales, el razonamiento “consecuencialista” (asociado aquí con procesos tipo 2) tiende a ser susceptible a efectos de carga cognitiva, mientras que el deontológico (asociado aquí con procesos tipo 1) no parece serlo (Greene, 2009; Greene, Morelli, Lowenberg, Nystrom y Cohen, 2008). Finalmente, estudios usando resonancia magnética

funcional sugieren que cada una de estas formas de evaluación moral están implementadas en diferentes estructuras neuronales: pensamiento “utilitarista” en la corteza prefrontal y los lóbulos parietales, zonas asociadas a la cognición; pensamiento “deontológico” en la amígdala y los lóbulos frontales y parietales, zonas asociadas a la emoción (Greene et al., 2001).

A pesar de su ubicuidad, la hipótesis general de los procesos duales ha sido cuestionada en tanto descripción adecuada de la arquitectura cognitiva humana. No es claro, en general, que los resultados conceptualizados de esta manera sean evidencia sólida de diferentes *tipos* de procesamiento cognitivo. Tampoco es claro, por razones conceptuales y empíricas, que tenga sentido hablar de dos procesos independientes determinantes del juicio moral, en vez de diferentes factores que interactúan entre sí (Una discusión de las críticas a la hipótesis general de sistemas duales pueden encontrarse en Keren y Schul (2009), Kruglanski y Gigerenzer (2011), Krajbich, Bartling, Hare y Fehr (2015) y Cohen (2017). Para escepticismo con respecto a su aplicación al caso del juicio moral, ver Helion y Pizarro (2014)).

Tal vez más importante para la presente discusión, la propuesta experimental de Greene y de sus colegas toma como herramienta central de trabajo dilemas sacrificiales, es decir, dilemas como el del tranvía en los que los participantes deben decidir si sacrifican a una persona para salvar varias vidas. Pero no es claro que este tipo de escenarios sean idóneos para el estudio del juicio moral. No solo estos incluyen situaciones que, por su artificialidad, ponen en tela de juicio la validez externa de las conclusiones obtenidas. Además de ello, estos dilemas parecen cubrir solo una pequeña parte del horizonte de posibles dilemas morales, los cuales no necesariamente implican contraposiciones entre formas de pensamiento deontológico y consecuencialista, sino entre tipos de valores y perspectivas evaluativas de otra naturaleza (al respecto, ver Rosas y Koenigs (2014), Bauman, McGraw, Bartels y Warren (2014) y Kahane (2015).

3. Juicios de responsabilidad

Junto con los juicios de estatus, los psicólogos morales también se han concentrado en lo que aquí llamamos juicios de responsabilidad moral. Aunque hay una variedad de posibles juicios de este tipo, el énfasis ha estado primariamente en juicios en los que a un agente se le atribuye responsabilidad por haber realizado *intencionalmente* una acción indebida o por haber omitido *intencionalmente* hacer algo frente a lo cual tenía algún tipo de obligación. En menor medida, el énfasis ha estado en atribuciones de responsabilidad por acciones buenas, no intencionales o en atribuciones de responsabilidad por rasgos de carácter y personalidad (ver, sin embargo, Janoff-Bulman (1979), Nadelhoffer (2004) y Lagnado y Channon (2008)).

Una forma común de aproximarse a la noción de responsabilidad moral, al menos desde una perspectiva filosófica tradicional, consiste en partir de la idea de que las personas son “autoras” de ciertos de sus comportamientos y de que en virtud de su autoría merecen diferentes tipos de reacciones como el elogio, la gratitud, la censura y el castigo (Strawson, 1962/2013; Wallace, 1994; Smith, 2005; McKenna, 2012). A modo de ilustración, estornudar, parpadear o tropezarse no suelen ser acciones adecuadas para las atribuciones de responsabilidad, pues uno no suele ser su autor. Uno es, más bien, pasivo ante ellas. Como tal, elogiar o censurar a alguien por llevar a cabo una de estas conductas no parece tener mucho sentido.

Desde luego, esta idea puede (y tal vez deba) ser problematizada. Algunos psicólogos morales, por ejemplo, han señalado que esta concepción de la responsabilidad se ve minada por el descubrimiento de que nuestras conductas morales parecen estar más guiadas por variables circunstanciales que por

disposiciones estables de carácter y/o personalidad (ver Doris (2002; 2015) y Vargas (2013b)). En su opinión, ello implica, cuando menos, una problematización de la noción de autoría, más probablemente, una re-conceptualización de la noción de responsabilidad en la que las estructuras sociales que la hacen posible jueguen un papel más central.

Poniendo a un lado este tipo de objeciones, una preocupación central para muchos filósofos morales ha sido la de explicar cuáles son las condiciones constitutivas de la responsabilidad (Feinberg, 1970; Fischer y Ravizza, 1998; Watson, 1996; Zimmerman, 1988). ¿Qué hace, en otras palabras, que una persona sea la “autora” de una determinada acción? ¿Cómo debe ser esta relación de la persona con sus acciones para que las reacciones de elogio y censura sean en principio justificables? En contraste, la psicología moral contemporánea se ha ocupado de una serie de preguntas ligeramente diferentes, a saber, no tanto en qué consiste la responsabilidad moral sino cuáles son las condiciones que suelen guiar nuestras atribuciones de responsabilidad (ver, no obstante, Malle y Knobe, 1997). ¿Cómo suele determinarse que una persona es la autora de sus comportamientos? ¿A qué tipo de factores suelen apelarse cuando se justifica el castigo y el elogio?

Como veremos a continuación, para algunos teóricos los juicios de responsabilidad o culpa cumplen una *función evaluativa* con respecto al sujeto al cual son dirigidos (Alicke, 2000; Uhlmann, Pizarro y Diermeier, 2015). Esto significa, en algunos casos, que las asignaciones de responsabilidad pueden entenderse como indicadores de la calidad moral de la voluntad o el carácter de la persona juzgada, tal como ésta es manifestada por sus acciones. En otras palabras, decir que alguien es culpable de algo implica afirmar que se trata de alguien que, al menos en virtud de esa acción, puede verse como malicioso, desatento o incluso moralmente torpe.

Para otros psicólogos morales, en cambio, los juicios de responsabilidad deben concebirse como el resultado de la aplicación de mecanismos generales de *razonamiento causal* a situaciones moralmente relevantes (Cushman y Young, 2011). Según esta corriente, las atribuciones de responsabilidad no están suscritas por evaluaciones globales de los sujetos juzgados, o de evaluaciones morales en general, sino que son diagnósticos de la estructura causal subyacente a su conducta (Sloman y Lagnado, 2015). Así, ser juzgado como responsable de una conducta o de sus consecuencias negativas es, puesto de otro modo, haber sido identificado como la causa (en un sentido que deber ser aclarado) de dicho comportamiento o de sus efectos.

En el resto de este capítulo, nos concentraremos en algunas de estas propuestas, destacando sus diferencias conceptuales y el tipo de evidencia empírica que las avala. Al respecto, cabe aclarar que la diferencia entre modelos evaluativos y causales que será discutida es más de grado de que principio. Ella radica, en últimas, en el tipo de consideraciones enfatizadas (consideraciones morales y normativas vs. consideraciones moralmente neutras sobre la estructura causal de la situación) que en una diferencia categórica. De igual forma, no es nuestra pretensión sugerir que esta es la única manera de clasificar las diferentes aproximaciones teóricas existentes. En efecto, otras categorizaciones de los modelos psicológicos de culpa y responsabilidad han sido formuladas (ver Alicke, Mandel, Hilton, Gerstenberg y Lagnado (2015) o Guglielmo (2015)).

3.1 Factores evaluativos

Una influencia importante en la psicología moral contemporánea proviene de las teorías de la atribución en psicología de mediados del siglo XX, representadas por teóricos como Heider (1958), Jones y Davis (1965) y Kelley (1973). De acuerdo con

esta tradición, las explicaciones cotidianas de los comportamientos humanos (lo que suele conocerse bajo el rótulo de *folk psychology*) están motivadas, no tanto por el deseo de comprender en abstracto por qué la gente se comporta como lo hace, sino porque dichas explicaciones nos permiten navegar socialmente e interactuar con otras personas (Krueger, 2007). Una motivación similar ha dado pie a una corriente en la investigación de los juicios de responsabilidad que hace énfasis en el aspecto motivacional de la culpa, es decir, en el propósito que las atribuciones de responsabilidad cumplen dentro de nuestra vida social como punto de inicio de la investigación sobre la etiología de los juicios de responsabilidad (ver, por ejemplo, Alicke (2000; 2014), Pizarro y Tannenbaum (2012) y Uhlmann et al. (2015)).

Mark Alicke (2000) ha desarrollado una propuesta influyente sobre este tipo de juicios de corte atribucionista. Su propuesta está inspirada en la idea de que la función principal de los juicios de responsabilidad es la de servir como mecanismos para hacer evaluaciones generales de las personas con las que interactuamos. De acuerdo con esta propuesta, los procesos de atribución de responsabilidad están guiados por reacciones afectivas que surgen primariamente de la valoración de comportamientos observados y que suelen transferirse a los individuos involucrados en ellos. Así, según Alicke, ante una circunstancia negativa relevante (una acción o sus consecuencias) la tendencia inicial predominante en las personas consiste en atribuir culpabilidad a los agentes involucrados, aun en ausencia de evidencia relativa a su autoría. De igual modo, información sobre la calidad moral de una persona suele potenciar, incluso en circunstancias de otro modo ambiguas, atribuciones de culpa (Alicke, 2014).

Alicke reconoce que, junto a estas consideraciones de corte evaluativo, hay también consideraciones sobre la autoría de las personas, enmarcadas en el control que éstas tienen sobre sus comportamientos: ¿es ésta una acción intencional?, ¿es la persona la causa primaria de su conducta?, ¿eran las consecuencias de ella

esperadas o previsibles? Aun así, su modelo asigna una cierta prioridad a las consideraciones evaluativas mencionadas anteriormente. Según éste, los procesos evaluativos preceden temporalmente a los procesos de determinación de control. Más importante aún, este tipo de diagnósticos de control personal a menudo está orientado a validar la atribución inicial de culpa. En el caso general, las consideraciones de control vienen a reforzar las atribuciones de culpa, en tanto que éstas están ya soportadas por consideraciones normativas.

Diferentes tipos de evidencia han sido presentados a favor de esta hipótesis. En primer lugar, existe evidencia de que, ante un evento negativo con varias causas potenciales, las personas tienden a elegir aquella que esté asociada a conductas o motivos reprochables (Alicke, 1992). El proceso de selección causal, en otras palabras, es sensible en estos casos de ambigüedad a consideraciones pragmáticas como el deseo de señalar o castigar a un agente culpable. De manera similar se ha observado que la confrontación con relatos de conductas indebidas aumenta las creencias tanto generales como específicas en el libre albedrío y que dicha relación parece estar mediada por la severidad del castigo “merecido” por los transgresores (Clark, Luguri, Ditto, Knobe, Shariff y Baumeister, 2014). Es decir, ante una trasgresión moral o evento negativo, las personas tienden a sobrestimar el control que una persona tuvo sobre su conducta con el fin de culparla y castigarla por ella (Alicke, 2000). Finalmente, se ha observado que las personas asignan mayores niveles de responsabilidad a un agente presentado como alguien indeseable que a un sujeto descrito como alguien socialmente atractivo (Alicke y Zell, 2009). En este sentido, la deseabilidad social del agente juzgado es una determinante importante de los juicios de responsabilidad.

3.2 Conceptos normativos

Si bien algunos psicólogos morales han enfatizado cómo las atribuciones de culpa suelen estar explícitamente determinadas por consideraciones normativas, otros han defendido una idea más radical. Según ésta, aspectos importantes de nuestras capacidades cognitivas asociadas con la interpretación de la conducta social de otros están, de entrada, permeadas por consideraciones morales. Como resultado de ello, las nociones psicológicas que conforman este tipo de juicios, es decir, nuestros conceptos de intencionalidad, causalidad y libertad, entre otros, tienen un carácter primariamente normativo y moral (Pettit y Knobe, 2009; Knobe, 2010; Phillips, Luguri y Knobe, 2015) que hace imposible distinguir claramente consideraciones evaluativas de consideraciones relativas al control que los agentes tienen de su conducta.

En este sentido, Joshua Knobe (2010) ha propuesto y articulado la hipótesis de la “moralidad dominante”. Su propuesta parte del hallazgo del “efecto del efecto colateral”, según el cual la valencia moral de los efectos secundarios de un determinado comportamiento incide en si dichos efectos son juzgados como intencionales o no. De acuerdo con los resultados obtenidos por el autor en una serie de estudios (Knobe, 2003a; 2003b; Knobe y Burra, 2006) efectos secundarios moralmente reprochables (e.g. dañar el medio ambiente) tienden a ser considerados intencionales con más probabilidad que efectos secundarios moralmente positivos (ayudar al medio ambiente). En su interpretación, esto sugiere que nuestra aplicación de conceptos como el de intencionalidad se ve previamente definida por consideraciones morales (Pettit y Knobe, 2009; Knobe, 2010).

Más interesante aún, esta asimetría no parece alterar solamente el proceso de atribución de intencionalidad. También influye en la determinación de contribuciones causales (Knobe y Fraser, 2008) y en un amplio rango de procesos relacionados, como asignar deseos y actitudes favorables a los agentes (Pettit y

Knobe, 2009) o determinar si una acción fue coaccionada o libremente escogida (Phillips y Knobe, 2009). A la luz de estos resultados Knobe (2010) propone que las consideraciones sobre la intención o la responsabilidad causal que preceden a los juicios de atribución de culpa dependen a su vez de otro tipo de juicio moral, a saber, sobre el *estatus moral* de la acción y de sus efectos colaterales. En este sentido, los conceptos como la intencionalidad o la causalidad, que sirven de antecedente a los juicios de culpa, no serían meramente descriptivos, sino que tendrían un carácter intrínsecamente moral (Phillips et al., 2015).

Aunque los resultados de Knobe con respecto al efecto colateral han sido ampliamente replicados, no por ello su propuesta ha estado libre de críticas. Por un lado, Nichols y Ulatowski (2007) advierten la importancia de tener en cuenta que una cantidad significativa de participantes en los experimentos de Knobe ofreció una respuesta contraria a la de la mayoría. Según los autores, esto sugiere que existen diferencias individuales en la comprensión del concepto de intencionalidad que deben ser reconocidas y explicadas. Paralelamente, Machery (2008) plantea que el concepto de acción intencional no es intrínsecamente normativo en virtud de que el “efecto del efecto colateral” se observa también en situaciones no-morales. Paradójicamente, fueron Knobe y Mendlow (2004) quienes advirtieron este hecho en un estudio en el cual los participantes debían determinar si el efecto secundario moralmente neutro (e.g. reducción de las ventas de una empresa en cierta ciudad) de una decisión había sido producido intencionalmente o no. Finalmente, Guglielmo y Malle (2010) han argumentado con base en evidencia empírica que las atribuciones de intencionalidad en los escenarios ideados por Knobe no dependen de la valencia moral del efecto colateral. Según los autores, dichas atribuciones dependen, primariamente, de ciertos estados mentales del agente juzgado, como sus deseos, que los participantes asignan con

base en la posible indiferencia de éste con respecto a ciertos de los efectos secundarios generados.

3.3 Consideraciones descriptivas

En los dos apartados previos examinamos modelos psicológicos según los cuales los juicios de responsabilidad están fuertemente influenciados (Alicke, 2000) o bien parecen estar constituidos (Knobe, 2010) por consideraciones normativas/evaluativas. Sin embargo, como mencionamos anteriormente, algunos psicólogos morales han defendido la tesis contraria de que los juicios de responsabilidad son el resultado de procesos generales de razonamiento causal y de atribución de estados mentales. Aunque estos modelos retoman algunos elementos centrales de las teorías de la atribución en psicología social (Heider, 1958; Shaver, 1985), lo esencial en ellos es que las asignaciones de responsabilidad surgen de procesos de razonamiento que no son exclusivos del dominio de la moralidad (Guglielmo, Monroe y Malle, 2009).

Un ejemplo ilustrativo de esta orientación puede encontrarse en el modelo propuesto por Bertram Malle y sus colegas (Malle et al., 2014). Computacionalmente hablando, éste consiste en un modelo de procesamiento jerárquico y secuencial en el cual la detección de un evento que viola alguna norma moral da pie a su examinación, primero, en términos de su dependencia causal con respecto a un agente y, segundo, con respecto a sus intenciones. Así, si se determina que el agente actuó intencionalmente, posteriormente se consideran las razones que podrían justificar su acción o, en caso de que se trate de una violación no intencional, se evalúan la posibilidad y la obligación que la persona tenía de actuar de otra manera.

En términos algorítmicos, cada etapa del modelo puede verse como un nodo conceptual en el que, una vez el concepto relevante se ha activado (mediante la detección inicial de un evento moralmente relevante o mediante la conclusión del procesamiento del concepto anterior), se inicia la búsqueda y adquisición de información específica para dicho concepto. En este sentido, por ejemplo, luego de constatar el vínculo causal entre un agente y un evento negativo, el concepto de intencionalidad se activa y se inicia un proceso de recolección de información concerniente a este criterio. Cuando la adquisición de información para un determinado concepto ha culminado satisfactoriamente (por ejemplo, si la información disponible indica que una transgresión fue intencional), se le asigna un valor (el valor correspondiente a “acción intencional”, por ejemplo) y se activa el siguiente concepto en la secuencia.

Una de las principales virtudes de este modelo es que su especificación precisa del tipo de procesos y contenidos involucrados en la asignación de responsabilidad permite formular hipótesis empíricamente evaluables. El modelo, por ejemplo, hace predicciones sobre qué tipo de información es normalmente requerida por participantes que han recibido historias ambiguas para hacer atribuciones de responsabilidad, así como sobre los tiempos de reacción relativos a la evaluación de diferentes historias. Varias de estas predicciones han sido examinadas y confirmadas experimentalmente (Guglielmo y Malle, 2017; Monroe y Malle, 2017).

Desde luego, este modelo también tiene ciertas limitaciones que no han pasado desapercibidas frente a sus críticos. Como algunos de ellos han señalado, esta propuesta ignora el rol que distintos factores, como la severidad de un resultado, el tipo de transgresión normativa cometida o el carácter moral de los agentes juzgados, tienen en las atribuciones de responsabilidad (Goodwin, 2014; Nadler, 2014; Uhlmann et al., 2015). Otros críticos han propuesto que los juicios de

responsabilidad surgen de la evaluación de un evento a la luz de un prototipo moral (compuesto por la presencia de un agente que infringe un daño a un paciente) y no a partir del examen secuencial de condiciones necesarias y suficientes para atribuir culpa (Schein y Gray, 2014). Finalmente, ciertos autores consideran que el modelo de trayectoria de la culpa es excesivamente racional en el sentido de que asume que las reacciones evaluativas, la motivación a culpar y las respuestas afectivas son elementos meramente secundarios, pero no constitutivos, de los juicios de responsabilidad (Alicke, 2014).

Aun cuando el modelo de trayectoria de la culpa es una de las propuestas existentes más completas, Malle y sus colaboradores ciertamente no han sido los únicos en otorgar primacía a consideraciones descriptivas en los juicios de atribución de culpabilidad. Otros autores han acumulado evidencia empírica de que los juicios de responsabilidad están mayormente basados en análisis de vinculación causal con consecuencias negativas (Cushman, 2008; Martin y Cushman, 2016). Incluso hay quienes sostienen una postura aún más radical, según la cual los modelos causales de un evento constituyen el principal tipo de representación cognitiva subyacente a los procesos de enjuiciamiento moral (Sloman, Fernbach y Ewing, 2009). En este tipo de perspectivas, las diferencias en la atribución de responsabilidad se reducen en últimas a la manera en la que los diferentes escenarios son construidos con relación a su estructura causal (Para algunas aplicaciones ver Lagnado y Gerstenberg (2017) y Wiegmann y Waldmann (2014)).

3.4 Invariantismo

Un presupuesto central en las teorías discutidas hasta ahora, tanto aquéllas de raigambre normativa como aquéllas que enfatizan los aspectos casuales del

razonamiento sobre responsabilidad, es que los juicios de responsabilidad provienen de una estructura cognitiva determinada (normativa o descriptiva) que emplea un conjunto estable de criterios en la formulación de esta clase de juicios. Este supuesto “invariantista”, sin embargo, ha sido puesto en duda, tanto por razones teóricas como por resultados empíricos (Doris, Knobe y Woolfolk, 2007). En general, el problema es que los juicios de responsabilidad parecen variar en función de la formulación (abstracta vs. concreta) del caso presentado, del estatus normativo (bueno vs. malo) de la conducta juzgada, de los antecedentes de los juicios de responsabilidad y de la cercanía entre quien juzga y el sujeto del juicio (Knobe y Doris, 2010).

Como ilustración, considérense los estudios empíricos de Malle y sus colegas (2014) mencionados anteriormente. Como vimos, en estos estudios se han utilizado primariamente viñetas que se refieren a instancias de daño personal (ver el material suplementario de Guglielmo y Malle, 2017; Monroe y Malle, 2017), cuya aplicación ha sugerido que la asignación de culpa sigue una lógica en la que el concepto de intencionalidad juega un papel central. Algunos autores, sin embargo, han planteado la posibilidad de que la centralidad de dicho concepto dependa fundamentalmente del dominio moral al que pertenece la conducta juzgada. Se ha visto, por ejemplo, que la intencionalidad parece ser un factor crítico a la hora de juzgar daños. Pero que éste juega un papel menos predominante cuando se trata de transgresiones de pureza, tales como el incesto. (Young y Saxe, 2011).

Estas observaciones pueden interpretarse de varias maneras. Una de ellas es que los resultados obtenidos en el estudio de los juicios de responsabilidad son, en gran parte, consecuencias de la manera en que las herramientas de estudio (viñetas y cuestionarios) han sido construidas. Otra interpretación más interesante es que lo que esta variabilidad muestra no es, simplemente, una limitación de nuestros estudios, sino la posibilidad de que *no haya una estructura cognitiva estable* que sea

común a todos los juicios de responsabilidad. Por el momento, no es claro cómo deba interpretarse esta disputa, pero presumiblemente ésta no podrá solucionarse sin un esfuerzo concertado por diversificar algunos de los materiales utilizados en el estudio de los juicios morales

4. Conclusión

En este capítulo nos hemos ocupado del juicio como una ventana hacia el estudio de la cognición moral. Para ello, hemos distinguido dos tipos de juicios morales: los juicios de estatus y los de responsabilidad. En la discusión hemos esbozado los principales debates en la literatura contemporánea sobre este tema. Con respecto a los juicios de estatus, hemos contrastado propuestas racionalistas y sentimentalistas. Con respecto a los juicios de responsabilidad, hemos contrastado modelos basados en consideraciones normativas/evaluativas y modelos basados en consideraciones causales/descriptivas.

Además de presentar estos debates, a lo largo de todo el texto nos interesó resaltar cómo la elección de los materiales experimentales ha influido en gran medida en el tipo de conclusiones a las que llegamos. Esto lo hemos visto tanto en el trabajo pionero de Kohlberg como en recientes modelos computacionales. De acuerdo con esto, creemos que los avances de los debates aquí abordados pueden provenir de un enriquecimiento y la diversificación del tipo de materiales y metodologías que usamos para estudiarlos.

Vale la pena aclarar que otras técnicas se han utilizado, aunque con menor frecuencia, en el estudio del juicio moral. Por ejemplo, algunos juegos económicos, como el paradigma del ultimátum y los experimentos de cooperación, se han usado para evaluar juicios de justicia (Pillutla y Murnighan, 1996) y atribuciones de castigo (Fehr y Gächter, 2002). Otros autores se han servido de las narrativas

autobiográficas para entender la culpa en primera persona (Baumeister, Stilwell, Heatherton, 1994) y el perdón (Zechmeister y Romero, 2002). Paralelamente, los juicios morales también se han estudiado a través de experimentos en los que se evalúan rutinas de búsqueda de información (Carlsmith, 2006; Guglielmo y Malle, 2017) y tiempos de reacción (Greene, Nystrom, Engell, Darley y Cohen, 2004; Schein y Gray, 2015, estudio 4). Además, aunque los estudios de resonancia magnética funcional han sido implementados con mucha frecuencia, otros indicadores fisiológicos, como la conductancia galvánica (Moretto, Ladavas, Mattioli y di Pellegrino, 2009) y los movimientos oculares o el diámetro de la pupila (Skulmowski, Bunge, Kaspar y Pipa, 2014; Kaspar, Krapp y Konig, 2015) también pueden usarse para evaluar ciertos aspectos de la cognición moral.

Finalmente, quisiéramos reiterar lo dicho inicialmente, con respecto al alcance de este capítulo. Si bien el estudio del juicio es una ventana hacia la cognición moral, el panorama de discusión en la psicología moral contemporánea es mucho más amplio que el que aquí hemos recorrido. Nuestra intención en el presente capítulo no ha sido tanto la de ser exhaustivos sino la de presentar un ejemplo de la riqueza de la investigación empírica en este campo.

Lecturas recomendadas

Las referencias incluidas en el cuerpo del artículo recogen una parte representativa de la discusión interdisciplinaria sobre el juicio moral en la psicología moral contemporánea. Para una visión general de esta última, el lector interesado puede consultar las siguientes antologías que reúnen a filósofos y científicos de la más diversa índole: Sinnott-Armstrong 2008 vols. 1-3, Doris et al. 2010, Nadelhoffer, Nahmias y Nichols 2010, D'Arms y Jacobson 2014 y Clausen y Levy 2015.

Además del juicio moral, otro tema central en la discusión contemporánea ha sido el de la motivación moral. Dentro de éste, el estudio de las *emociones morales* ha tenido una importancia que vale la pena destacar. Pues si bien, como hemos visto, éstas pueden verse como expresiones de evaluaciones morales, ellas también son motivadoras importantes de comportamientos pro-sociales (Fiske, 2002; Hardy, 2006; Hutcherson y Gross, 2011; Keltner y Haidt, 1999; Schroeder, Roskies y Nichols, 2010; Teper, Inzlicht, y Page-Gould, 2011).

Dentro de las emociones morales, las emociones “negativas” han recibido mayoritariamente la atención. El orgullo (Carter y Gordon, 2017), la rabia (Cherry y Flanagan, 2017), la culpa (Baumeister, Stillwell y Heatherton 1994), el asco (Kelly, 2011; Strohminger y Kumar, 2018) y el desprecio (Mason, 2018) son ejemplos de algunas emociones estudiadas a fondo. También, emociones positivas como la compasión (Caouette y Price, 2018; Goetz, Keltner, y Simon-Thomas, 2010), la gratitud (Algoe y Haidt, 2009; DeSteno, Condon y Dickens, 2016), la admiración (Delin y Baumeister, 1994; Keltner y Haidt, 2003), el amor (Abramson y Leite, 2011; Fredrickson, 2016; Kolodny, 2003) y la inspiración moral (Schnall, Roper y Fessler, 2010; Silvers y Haidt, 2008) han sido estudiadas sistemáticamente.

Una segunda vertiente de la discusión sobre motivación moral tiene que ver con el debate entre *egoísmo vs. altruismo*. Desde una perspectiva biológica, la pregunta ha sido por la evolución de conductas que promuevan el bienestar de otros individuos a costa del beneficio propio. Tradicionalmente, los defensores de la teoría de selección de parentesco han argumentado que las conductas altruistas están mediadas por relaciones de cercanía genética (ver la formulación clásica de esta idea en Hamilton, 1964 y una defensa reciente en Foster, Wenseleers y Ratnieks, 2006). En contraste, otros teóricos han sostenido que el altruismo surge también en relaciones de cooperación a largo plazo (Trivers 1971; 2006; Axelrod, 1984; 2000) e incluso que, en ciertos contextos, motivaciones altruistas pueden

conferir ventajas evolutivas (una exposición clásica de esta idea puede encontrarse en Sober y Wilson (1998), aunque véanse también las críticas de Rosas, 2002 y Stich, 2007). Es posible, sin embargo, que la disputa entre estas posiciones se deba, no tanto a desacuerdos sustantivos, como a diferencias en sus niveles de análisis (Fletcher y Zwick, 2006; Kurzban, Burton-Chellew y West, 2015).

Por el contrario, desde una perspectiva social, la pregunta se ha concentrado en los mecanismos psicológicos subyacentes al altruismo. Notablemente, Daniel Batson (1991; 2011) ha acumulado evidencia empírica según la cual la empatía genera comportamientos altruistas que no pueden explicarse por motivos egoístas, como evitar un castigo (Batson et al., 1988), obtener una recompensa (Batson et al., 1991) o disminuir la exposición al dolor ajeno (Batson et al., 1983). En respuesta, algunos autores han propuesto interpretaciones alternativas de estos resultados (Cialdini, 1991; Cialdini et al., 1997). Otros han cuestionado el constructo teórico de la empatía a la base de estos (Maibom 20XX). Más recientemente, la deseabilidad moral de la empatía como mecanismo de motivación ha sido cuestionada (al respecto, véase Prinz 2011, Bloom 2017 y los ensayos en Coplan y Goldie 2011)

Un tercer tema relacionado con la motivación moral tiene que ver con el papel que *el carácter y la personalidad* juegan en el comportamiento moral de las personas. Al respecto, varios teóricos han cuestionado la idea de que éste pueda explicarse en términos de la posesión de ciertos rasgos positivos de carácter (véase Harman, 2000; Doris, 2002; Alfano, 2013). Sus argumentos, basados una serie de resultados en psicología social, han estado dirigidos a mostrar que los rasgos comúnmente invocados en este contexto (valentía, lealtad, etc.) no tienen la estabilidad o la consistencia requerida para superar presiones situacionales contrarias. Y, con base en esto, han criticado a la llamada ética de las virtudes.

Obviamente, tratándose de una venerable tradición filosófica, varias voces han respondido a este “escepticismo” sobre el carácter. Algunas se han opuesto a

éste apelando la idea de que las virtudes, en cuanto nociones reguladores e ideales, no pueden por principio ser cuestionadas por este tipo de evidencia (Badhwar, 2009; Kupperman, 2009). Otros teóricos han argumentado que la ética de las virtudes puede formularse con independencia de algunas de las presuposiciones psicológicas allí cuestionadas (Merritt, 2000; Miller, 2009). Finalmente, algunos teóricos han aceptado la relevancia de la evidencia empírica en general, pero han cuestionado que esta evidencia tenga la fuerza aducida por los escépticos. (Kamtekar, 2004; Russell, 2009). Sea como fuere, dicha discusión vale la pena de ser estudiada pues, más allá de sus particularidades, ha ayudado a establecer la relevancia de consideraciones empíricas en el desarrollo de teorías normativas en filosofía (Al respecto de esto último ver también Doris et al., 2010 y D'Arms y Jacobson, 2014).

Referencias

- Abramson, K. y Leite, A. (2011). Love as a Reactive Emotion. *Philosophical Quarterly* 61 (245), 673-699.
- Alfano, M. (2013). Identifying and Defending the Hard Core of Virtue Ethics. *Journal of Philosophical Research*, 38, 233-260.
- Algoe, S. B. y Haidt, J. (2009). Witnessing excellence in action: The “otherpraising” emotions of elevation, gratitude, and admiration. *Journal of Positive Psychology*, 4 (2), 105–127.
- Alicke, M. D. (1992). Culpable causation. *Journal of personality and social psychology*, 63 (3), 368-378.
- Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological bulletin*, 126 (4), 556-574.
- Alicke, M. D. (2014). Evaluating blame hypotheses. *Psychological Inquiry*, 25 (2), 187-192.
- Alicke, M. D. y Zell, E. (2009). Social attractiveness and blame. *Journal of Applied Social Psychology*, 39 (9), 2089-2105.
- Alicke, M. D., Mandel, D. R., Hilton, D. J., Gerstenberg, T. y Lagnado, D. A. (2015). Causal conceptions in social explanation and moral evaluation: A historical tour. *Perspectives on Psychological Science*, 10 (6), 790-812.
- Anscombe, G. (1958). Modern moral philosophy. *Philosophy*, 33 (124), 1-19.
- Axelrod, R. M. (2000). On six advances in cooperation theory. *Analyse & Kritik*, 22(1), 130-151.

- Axelrod, R.M. (1984). *The Evolution of Cooperation*. Nueva York: Basic Books.
- Badhwar, N. K. (2009). The Milgram Experiments, Learned Helplessness, and Character Traits. *The Journal of Ethics*, 13(2-3), 257-289.
- Barbosa, S. y Jiménez, W. (2017). It's not right but it's permitted: Wording effects in moral judgement. *Judgment and Decision Making*, 12 (3), 308-313.
- Batson, C. D. (1991). *The Altruism Question: Toward a Social-Psychological Answer*, Hillsdale, NJ: Lawrence Erlbaum Associates.
- Batson, C. D. (2011), *Altruism in Humans*, Oxford: Oxford University Press.
- Batson, C. D., Batson, J. G., Slingsby, J. K., Harrell, K. L., Peekna, H. M., & Todd, R. M. (1991). Empathic joy and the empathy-altruism hypothesis. *Journal of personality and social psychology*, 61(3), 413-426.
- Batson, C. D., Dyck, J. L., Brandt, J. R., Batson, J. G., Powell, A. L., McMaster, M. R., & Griffitt, C. (1988). Five studies testing two new egoistic alternatives to the empathy-altruism hypothesis. *Journal of personality and social psychology*, 55(1), 52-77.
- Batson, C. D., O'Quin, K., Fultz, J., Vanderplas, M., & Isen, A. M. (1983). Influence of self-reported distress and empathy on egoistic versus altruistic motivation to help. *Journal of personality and social psychology*, 45(3), 706-718.
- Bauman, C. W., McGraw, A. P., Bartels, D. M. y Warren, C. (2014). Revisiting External Validity: Concerns about Trolley Problems and Other Sacrificial Dilemmas in Moral Psychology. *Social and Personality Psychology Compass*, 8/9, 536-554.
- Baumeister, R. F., Stillwell, A. M., y Heatherton, T. F. (1994). Guilt: An interpersonal approach. *Psychological Bulletin*, 115 (2), 243-267.
- Bechara, A., Damasio, H. y Damasio, A.R. (2000). Emotion, decision-making and the orbitofrontal cortex. *Cerebral Cortex*, 10, 295-307.
- Bjorklund, F. (2003). Differences in the justification of choices in moral dilemmas: Effects of gender, time pressure and dilemma seriousness. *Scandinavian Journal of Psychology*, 44, 459-466.
- Blair, R. (1993). *The development of morality*. Londres: University of London.
- Blair, R. (1995). A cognitive developmental approach to morality: investigating the psychopath. *Cognition*, 57, 1-29.
- Blair, R. (1997). Moral Reasoning and the Child with Psychopathic Tendencies. *Personality and Individual Differences*, 22, 731-739.
- Blair, R. y Cipolotti, L. (2000). Impaired social response reversal: A case of 'acquired sociopathy'. *Brain*, 123, 1122-1141.
- Bloom, P. (2017). Empathy and its discontents. *Trends in Cognitive Sciences*, 21(1), 24-31.
- Caouette, J. y Price, C. (Eds.). (2018). *The Moral Psychology of Compassion*. Londres: Rowman & Littlefield International.
- Carter, J. A. y Gordon E. C. (Eds.) (2017). *The Moral Psychology of Pride*. Londres: Rowman & Littlefield International.
- Cherry, M. y Flanagan, O. (Eds.) (2017). *The Moral Psychology of Anger*. Londres: Rowman & Littlefield International.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Nueva York: Praeger.

- Cialdini, R. B. (1991). Altruism or egoism? That is (still) the question. *Psychological Inquiry*, 2(2), 124-126.
- Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C., & Neuberg, S. L. (1997). Reinterpreting the empathy–altruism relationship: When one into one equals oneness. *Journal of personality and social psychology*, 73(3), 481-494.
- Clark, C. J., Luguri, J. B., Ditto, P. H., Knobe, J., Shariff, A. F. y Baumeister, R. F. (2014). Free to punish: a motivated account of free will belief. *Journal of personality and social psychology*, 106 (4), 501-513.
- Clausen, J., & Levy, N. (Eds.). (2015). *Handbook of Neuroethics*. Dordrecht: Springer.
- Cohen, J.D. (2017). Cognitive Control: Core Constructs and Current Considerations. En T. Egner (Ed.), *Wiley Handbook of Cognitive Control* (pp. 3-28). New York: Wiley Blackwell.
- Colby, A., Kohlberg, L., Gibbs, J., Lieberman, M., Fischer, K. y Saltzstein, H. D. (1983). A longitudinal study of Moral Judgment. *Monographs of the Society for Research in Child Development*, 48 (1-2, Serial No. 200) (pp. 1-124). Chicago: University of Chicago Press.
- Coplan, A. y Goldie, P (Eds.) (2011), *Empathy: Philosophical and Psychological Perspectives*. Oxford: Oxford University Press.
- Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, 108 (2), 353-380.
- Cushman, F. y Young, L. (2011). Patterns of moral judgment derive from nonmoral psychological representations. *Cognitive Science*, 35 (6), 1052-1075.
- Cushman, F., Young, L. y Greene, J. D. (2010). Our multi-system moral psychology: Towards a consensus view. *The Moral Psychology Handbook*. Nueva York: Oxford University Press.
- Cushman, F., Young, L. y Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment testing three principles of harm. *Psychological Science*, 17 (12), 1082–1089.
- D'Arms, J. (2005). Two Arguments for Sentimentalism. *Philosophical Issues*, 15 (1), 1-21.
- Damasio, A.R. (1995). *Descartes' error: emotion, reason and the human brain*. Nueva York: Avon Books.
- Daniels, N. (1979). Wide reflective equilibrium and theory acceptance in ethics. *The Journal of Philosophy*, 76 (5), 256-282.
- D'Arms, J., y Jacobson, D. (2014). *Moral Psychology and Human Agency: Philosophical Essays on the Science of Ethics*. Oxford: Oxford University Press.
- Delin, C. R. y Baumeister, R. F. (1994). Praise: More than just social reinforcement. *Journal for the Theory of Social Behaviour*, 24 (3), 219-241.
- DeSteno, D., Condon, P. y Dickens, L. (2016). Gratitude and Compassion. En L. F. Barrett, M. Lewis y J. M. Haviland (Eds.), *Handbook of Emotions*, (pp. 847-858). Nueva York, NY: Guilford Press.
- Doris, J. M. (2002). *Lack of character: Personality and moral behavior*. New York: Cambridge University Press.
- Doris, J. M. y The Moral Psychology Research Group. (2010). *The moral psychology handbook*. Oxford: Oxford University Press.

- Doris, J. M., Knobe, J. y Woolfolk, R. L. (2007). Variantism about responsibility. *Philosophical perspectives*, 21 (1), 183-214.
- Doris, J.M. (2015). *Talking to Our Selves: Reflection, ignorance and agency*. Oxford: Oxford University Press.
- Doris, J.M. y Stich, S. (2014). Moral Psychology: Empirical Approaches. En E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Recuperado de <<https://plato.stanford.edu/archives/fall2014/entries/moral-psych-emp/>>.
- Driver, J. (2007). *Ethics: The Fundamentals*. Oxford: Blackwell.
- Dunn, J. y Munn, P. (1987). Development of justification in disputes with mother and sibling. *Developmental Psychology*, 23, 791–98.
- Dwyer, S. (1999). Moral competence. En K. Murasugi y R. Stainton (Eds.), *Philosophy and linguistics* (pp. 169–190). Boulder, CO: Westview Press.
- Dwyer, S. (2006). How good is the linguistic analogy? En P. Carruthers, S. Laurence y S. P. Stich (Eds.), *The Innate Mind, Vol. 2: Culture and Cognition* (pp. 145-167). Nueva York: Oxford University Press.
- Eisenberg, N. (1979). Development of children's prosocial moral judgment. *Developmental Psychology*, 15 (2), 128-137.
- Eisenberg, N. (1986). *Altruistic Emotion, Cognition, and Behavior (Child Psychology)*. Nueva Jersey: Lawrence Erlbaum.
- Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences*. Brighton, England: Erlbaum.
- Evans, J. St. B. T. y Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, 8 (223), 223-241.
- Fehr, F. y Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868), 137-140.
- Feinberg, J. (1970). *Doing and Deserving: Essays in the Theory of Responsibility*. Princeton: Princeton University Press.
- Fischer, J. M. y Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.
- Fiske, A. P. (2002). Moral emotions provide the self-control needed to sustain social relationships. *Self and Identity*, 1, 169 –175.
- Fletcher, J. A., & Zwick, M. (2006). Unifying the theories of inclusive fitness and reciprocal altruism. *The American Naturalist*, 168(2), 252-262.
- Foot, P. (1967). The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review*, 5, 5–15.
- Foster, K. R., Wenseleers, T., & Ratnieks, F. L. (2006). Kin selection is the key to altruism. *Trends in Ecology & Evolution*, 21(2), 57-60.
- Fredrickson, B. L. (2016). Love: Positivity resonance as a fresh, evidence-based perspective on an age-old topic. En L. F. Barrett, M. Lewis y J. M. Haviland (Eds.), *Handbook of Emotions*, (pp. 847-858). Nueva York, NY: Guilford Press.
- Fultz, J., Batson, C. D., Fortenbach, V. A., McCarthy, P. M., & Varney, L. L. (1986). Social evaluation and the empathy–altruism hypothesis. *Journal of personality and social psychology*, 50(4), 761.

- Gilligan, C. (1982). *In a different voice: psychological theory and women's development*. Cambridge, MA: Harvard University Press.
- Gilligan, C. (1995). Hearing the Difference: Theorizing Connection. *Hypatia*, 10 (2), 120-127.
- Goetz, J. L., Keltner, D. y Simon-Thomas, E. (2010). Compassion: An evolutionary analysis and empirical review. *Psychological Bulletin*, 136 (3), 351–374.
- Goodman, N. (1955). *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.
- Goodwin, G. P. (2014). How Complete Is the Path Model of Blame? *Psychological Inquiry*, 25 (2), 215-221.
- Greene, J. D. (2008). The secret joke of Kant's soul. En W. Sinnott-Armstrong (Ed.), *Moral Psychology: The Neuroscience of Morality: Emotion, Disease, and Development (Vol. 3)* (pp. 35-79). Cambridge, MA: MIT Press.
- Greene, J. D. (2009). Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology* 45 (3), 581-584.
- Greene, J. D. (2013). *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. Nueva York: Penguin Press.
- Greene, J. D. y Haidt, J. (2002). How (and where) does moral judgment work? *TRENDS in Cognitive Sciences*, 6 (12), 517-523.
- Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E. y Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition* 107, 1144–1154.
- Greene, J. D., Sommerville, B. R., Nystrom, L. E., Darley, J. M. y Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293, 2105-2018.
- Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M. y Cohen, J.D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44(2), 389-400.
- Gruber, J., Mauss, I. B. y Tamir, M. (2011). A dark side of happiness? How, when, and why happiness is not always good. *Perspectives on Psychological Science* 6(3), 222 – 233.
- Guglielmo, S. (2015). Moral judgment as information processing: an integrative review. *Frontiers in Psychology*, 6, 1-19.
- Guglielmo, S. y Malle, B. F. (2010). Can unintended side effects be intentional? Resolving a controversy over intentionality and morality. *Personality and social psychology bulletin*, 36 (12), 1635-1647.
- Guglielmo, S. y Malle, B. F. (2017). Information-Acquisition Processes in Moral Judgments of Blame. *Personality and Social Psychology Bulletin*, 43 (7), 957-971.
- Guglielmo, S., Monroe, A. E. y Malle, B. F. (2009). At the heart of morality lies folk psychology. *Inquiry*, 52 (5), 449-466.
- Haidt, J. (2001). The emotional dog and its rational tail. *Psychological Review*, 108, 814-834.
- Haidt, J. (2007). The New Synthesis in Moral Psychology. *Science*, 316 (5827), 998-1002.
- Haidt, J., Björklund, F., & Murphy, S. (2000). Moral dumbfounding: When intuition finds no reason. Manuscript sin publicar, Universidad de Virginia.
- Haidt, J., Koller, S. y Dias, M. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65, 613-628.
- Hamilton, W.D. (1964) The genetical evolution of social behaviour. I & II. *Journal of Theoretical Biology*, 7, 1–52

- Hardy, S. A. (2006). Identity, reasoning, and emotion: An empirical comparison of three sources of moral motivation. *Motivation and Emotion*, 30 (3), 207–215.
- Harman, G. (1999). Moral philosophy and linguistics. En K. Brinkmann (Ed.), *Proceedings of the 20th World Congress of Philosophy: Vol. 1. Ethics* (pp. 107–115). Bowling Green, OH: Philosophy Documentation Center.
- Harman, G. (2000). The nonexistence of character traits. *Proceedings of the Aristotelian Society*, 100(2), 223–226.
- Hauser, M. D., Cushman, F., Young, L., Jin, R. K. X. y Mikhail, J. (2006). A dissociation between moral judgment and justifications. *Mind & Language*, 22, 1–21.
- Hauser, M., Tonnaer, F. y Cima, M. (2010). Psychopaths know right from wrong but don't care. *Social Cognitive Affective Neuroscience*, 5 (1), 59-67.
- Hauser, M., Young, L. y Cushman, F. (2008). Reviving Rawls's Linguistic Analogy: Operative Principles and the Causal Structure of Moral Actions. En W. Sinnott Armstrong (Ed.), *Moral Psychology. Volume 2. The Cognitive Science of Morality: Intuition and Diversity* (pp. 107-143). Cambridge: MIT Press.
- Heider, F. (1958). *The Psychology of Interpersonal Relations*. Hoboken: Wiley.
- Helion, C. y Pizarro, D.A. (2015). Beyond dual-processes: The interplay of reason and emotion in moral judgment. En N. Levy y J. Clausen (Eds.), *Springer Handbook for Neuroethics* (pp 109-125). Dordrecht: Springer.
- Huebner, B., Lee, J. y Hauser, M. (2010). The Moral-Conventional Distinction in Mature Moral Competence. *Journal of Cognition and Culture*, 10 (1-2), 1-26.
- Hume, D. (1739/2005). *Tratado de la naturaleza humana*. Madrid: Tecnos.
- Hume, D. (1751/2006). *Investigación sobre los principios de la moral*. Madrid: Alianza.
- Hutcherson, C. A. y Gross, J. J. (2011). The moral emotions: a social-functionalist account of anger, disgust, and contempt. *Journal of Personality and Social Psychology* 100 (4), 719-37.
- Janoff-Bulman, R. (1979). Characterological versus behavioral self-blame: inquiries into depression and rape. *Journal of personality and social psychology*, 37 (10), 1798-1809.
- Jones, E. E. y Davis, K. E. (1965). From acts to dispositions. The attribution process in person perception. *Advances in Experimental Social Psychology*, 2, 219-266.
- Kahane, G. (2015). Sidetracked by Trolleys. Why sacrificial moral dilemmas tell us little (or nothing) about utilitarian judgment. *Social Neuroscience*, 10 (5), 551/560.
- Kahneman, D. (2011). *Thinking, fast and slow*. Nueva York, NY: Farrar, Straus and Giroux.
- Kamtekar, R. (2004). Situationism and virtue ethics on the content of our character. *Ethics*, 114(3), 458-491.
- Kant, E. (1781/2011). *Crítica de la razón pura*. (Traducción de Mario Caimi). México D.F.: Fondo de Cultura Económica.
- Kant, E. (1785/2002). *Fundamentación para una metafísica de las costumbres*. Madrid: Alianza.
- Kaspar, K., Krapp, V. Y Konig, P. (2015). Hand washing induces a clean slate effect in moral judgments: A pupillometry and eye-tracking study. *Scientific Reports*, 5, 1-10.
- Kelley, H. H. (1973). The processes of causal attribution. *American psychologist*, 28 (2), 107-128.

- Kelly, D. (2011). *Yuck!: The Nature and Moral Significance of Disgust (Life and Mind: Philosophical Issues in Biology and Psychology)*. Cambridge, MA: MIT Press.
- Kelly, D., Stich, S., Haley, K., Eng, S. J. y Fessler D. M. T. (2007). Harm, Affect, and the Moral/Conventional Distinction. *Mind & Language*, 22 (2), 117-131.
- Keltner, D. y Haidt, J. (1999). Social functions of emotion at four levels of analysis. *Cognition & Emotion*, 13 (5), 505-521.
- Keltner, D. y Haidt, J. (2003). Approaching awe, a moral, spiritual, and aesthetic emotion. *Cognition & Emotion*, 17 (2), 297-314.
- Keren, G. y Shul, Y. (2009). Two is not always better than one: acritical evaluation of two systems theory. *Perspectives on Psychological Science*, 4, 533-550.
- Kinga, J., Blair, R., Derek, R., Mitchell, G., Dolan, R. y Burgess, N. (2005). Doing the right thing: A common neural circuit for appropriate violent or compassionate behavior. *Neuroimage*, 30, 1069-1076.
- Knobe, J. (2003a). Intentional action and side effects in ordinary language. *Analysis*, 63 (279), 190-194.
- Knobe, J. (2003b). Intentional action in folk psychology: An experimental investigation. *Philosophical psychology*, 16 (2), 309-324.
- Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences*, 33 (4), 315-329.
- Knobe, J. y Burra, A. (2006). Intention and Intentional Action: A Cross-Cultural Study. *Journal of Culture and Cognition*, 6, 113-132.
- Knobe, J. y Doris, J. (2010). Responsibility. En J. Doris (Ed.), *The handbook of moral psychology* (pp. 321-353). Oxford: Oxford University Press.
- Knobe, J. y Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. En W. Sinnott-Armstrong (Ed.), *Moral psychology (Vol. 2): The cognitive science of morality: Intuition and diversity* (pp. 441-447). Cambridge: MIT Press.
- Knobe, J. y Mendlow, G. S. (2004). The Good, the Bad and the Blameworthy: Understanding the Role of Evaluative Reasoning in Folk Psychology. *Journal of Theoretical and Philosophical Psychology*, 24 (2), 252-258.
- Koenigs, M., Kruepke, M., Zeier, J. y Newman, J. P. (2012). Utilitarian moral judgment in psychopathy. *Social Cognitive and Affective Neuroscience*, 7 (6), 708-14.
- Kohlberg, L. (1984/1992). *Psicología del desarrollo moral*. España: Editorial Desclée de Broouwer.
- Kohlberg, L. y Hersh, R. (1977). Moral development: A review of the theory. *Theory Into Practice*, 16 (2), 53-59.
- Kolodny, N. (2003). Love as a Valuing Relationship. *Philosophical Review* 112 (2), 135-189.
- Krajbich, I., Bartling, B., Hare, T., Fehr, E. (2015). Rethinking fast and slow based on a critique of a reaction-time reversed inference. *Nature Communications*, 2 (6), 7745.
- Krueger, J. (2007). Attribution Theory. En K.D. Vohs y R.F. Baumeister (Eds.), *Encyclopedia of Social Psychology* (pp. 75-78). Thousand Oaks: Sage Publications.
- Kruglanski, A. y Gigerenzer, G. (2011) Intuitive and deliberative judgments are based on common principles. *Psychological Review*, 118, 97-109.
- Kupperman, J. J. (2009). Virtue in Virtue Ethics. *The Journal of Ethics*, 13(2-3), 243-255.

- Kurzban, R., Burton-Chellew, M. N., & West, S. A. (2015). The evolution of altruism in humans. *Annual Review of Psychology*, 66, 575-599.
- Lagnado, D. A. y Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, 108 (3), 754-770.
- Lagnado, D. A. y Gerstenberg, T. (2017). Causation in legal and moral reasoning. En M. Waldmann (Ed.), *The Oxford Handbook of Causal Reasoning* (pp. 565-601). Oxford: Oxford University Press.
- Lapsley, D. (2005). Moral stage theory. En M. Killen y J. Smetana (Eds.), *Handbook of moral development* (pp. 37-66). Mahwah, NJ: Lawrence Erlbaum Associates.
- Lapsley, D. K. (1996). *Moral psychology*. Boulder: Westview Press.
- Lapsley, D. y Narvaez, D. (2005). Moral psychology at the crossroads. En D. Lapsley y C. Power (Eds.), *Character Psychology and Character Education* (pp. 18-35). Notre Dame: University of Notre Dame Press.
- Levy, N. (2005). Imaginative Resistance and the Moral/Conventional Distinction. *Philosophical Psychology*, 18 (2), 231 – 241.
- Machery, E. (2008). The folk concept of intentional action: Philosophical and experimental issues. *Mind & Language*, 23 (2), 165-189.
- Maibom, H. (2010) Imagining Others. *Les Ateliers de l'éthique*, 5(1), 34-49.
- Malle, B. F., Guglielmo, S. y Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry*, 25 (2), 147-186.
- Malle, B.F. y Knobe, J. (1997). The folk concept of intentional action. *Journal of Experimental Social Psychology*, 33(2), 101-121.
- Martin, J. W. y Cushman, F. (2016). Why we forgive what can't be controlled. *Cognition*, 147, 133-143.
- Mason, M. (Ed.) (2018). *The Moral Psychology of Contempt*. London: Rowman & Littlefield International.
- McIntyre, A. (2014). Doctrine of Double Effect. En E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Recuperado de <<https://plato.stanford.edu/archives/win2014/entries/double-effect/>>.
- McKenna, M. (2012). *Conversation and responsibility*. New York: Oxford University Press.
- Mendez, M., Anderson, E. y Shapira, J. (2005). An investigation of moral judgment in frontotemporal dementia. *Cognitive Behavioral Neurology*, 18, 193-197.
- Merritt, M. (2000). Virtue ethics and situationist personality psychology. *Ethical Theory and Moral Practice*, 3(4), 365-383.
- Mikhail, J. (2000). Rawls' linguistic analogy: A study of the "generative grammar" model of moral theory described by John Rawls in "A theory of justice." Ithaca: Cornell University.
- Mikhail, J. (2002). Aspects of the Theory of Moral Cognition: Investigating Intuitive Knowledge of the Prohibition of Intentional Battery and the Principle of Double Effect. *Georgetown Law and Economics Research Paper No. 762385*.
- Miller, C. (2009). Empathy, social psychology, and global helping traits. *Philosophical Studies*, 142(2), 247-275.

- Monroe, A. E. y Malle, B. F. (2017). Two paths to blame: Intentionality directs moral information processing along two distinct tracks. *Journal of Experimental Psychology*, 146 (1), 123-133.
- Moore, G.E. (1903/1993). *Principia Ethica*. Cambridge: Cambridge University Press.
- Moretto, G., Ladavas, E., Mattioli, F. y di Pellegrino, G. (2010). A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *Journal of cognitive neuroscience*, 22(8), 1888-1899.
- Myers, D. G. y Diener, E. (1995). Who is happy? *Psychological Science* 6(1), pp. 10 – 19.
- Nadelhoffer, T. (2004). On Praise, Side Effects, and Folk Ascriptions of Intentionality. *Journal of Theoretical and Philosophical Psychology*, 24 (2), 196-213.
- Nadelhoffer, T., Nahmias, E., & Nichols, S. (Eds.). (2010). *Moral psychology: historical and contemporary readings*. Hoboken: John Wiley & Sons.
- Nadler, J. (2014). The Path of Motivated Blame and the Complexities of Intent. *Psychological Inquiry*, 25 (2), 222-229.
- Nahmias, E., Morris, S. G., Nadelhoffer T. y Turner, J. (2005). Surveying Freedom: Folk Intuitions about free will and moral responsibility. *Philosophical Psychology* 18 (5), 561-584.
- Nichols, S. (2002). How psychopaths threaten moral rationalism: is it irrational to be amoral? *The Monist*, 85, 285–304.
- Nichols, S. (2002). Norms with feeling: Towards a psychological account of judgment. *Cognition*, 84, 221–36.
- Nichols, S. (2014). Process debunking and ethics. *Ethics*. 124 (4), 727-749.
- Nichols, S. y Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs*, 41 (4), 663–685.
- Nichols, S. y Ulatowski, J. (2007). Intuitions and individual differences: The Knobe effect revisited. *Mind & Language*, 22 (4), 346-365.
- Nichols, S. (2004). *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Nueva York: Oxford University Press.
- Nisan, M. (1987). Moral norms and social conventions: A cross-cultural comparison. *Developmental Psychology*, 23, 719-725.
- Nucci, L. (1986). Children's conceptions of morality, social conventions and religious prescription. En C. Harding (Ed.), *Moral dilemmas: Philosophical and psychological reconsiderations of the development of moral reasoning*. Chicago: Precedent Press.
- Nucci, L. P. y Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. *Child Development*, 49 (2), 400–407.
- Nucci, L. P. y Turiel, E. (1993). God's word, religious rules, and their relation to Christian and Jewish children's concepts of morality. *Child Development*, 64, 1475-1491.
- O'Hara, R. E., Sinnott-Armstrong, W. y Sinnott-Armstrong, N. A. (2010). Wording effects in moral judgment. *Judgment and Decision Making*, 5 (7), 547–554.
- O'Neill, E. (2017). Kinds of norms. *Philosophy Compass*, 12 (5), 1-15.
- Pettit, D. y Knobe, J. (2009). The pervasive impact of moral judgment. *Mind & Language*, 24 (5), 586-604.

- Phillips, J. y Knobe, J. (2009). Moral judgments and intuitions about freedom. *Psychological Inquiry*, 20, 30-36.
- Phillips, J., Luguri, J. B. y Knobe, J. (2015). Unifying morality's influence on non-moral judgments: The relevance of alternative possibilities. *Cognition*, 145, 30-42.
- Piaget, J. (1932/1965). *The Moral Judgment of the Child*. New York: Free Press.
- Pillutla, M.M. y Murnighan, J.K. (1996). Unfairness, Anger and Spite: Emotional Rejections of Ultimatum Offers. *Organizational Behavior and Human Decision Processes*. 68(3), 208-224).
- Pizarro, D. A. y Tannenbaum, D. (2012). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. En M. Mikulincer y P. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 91-108). Washington, D.C.: American Psychological Association.
- Prinz, J. (2006). The emotional basis of moral judgments. *Philosophical Explorations*, 9 (1), 29-43.
- Prinz, J. (2008). *The Emotional Construction of Morals*. Nueva York: Oxford University Press.
- Prinz, J. (2011). Is empathy necessary for morality. In Amy Coplan & Peter Goldie (eds.), *Empathy: Philosophical and Psychological Perspectives*. Oxford University Press. pp. 211--229 (2011)
- Rawls, J. (1955). Two concepts of rules. *Philosophical Review*, 64 (1), 3-32.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Rosas, A. (2002). Psychological and evolutionary evidence for altruism. *Biology and Philosophy*, 17(1), 93-107.
- Rosas, A. y Koenigs, M. (2014). Beyond "utilitarianism": maximizing the clinical impact of moral judgment research. *Social Neuroscience* 9 (6), 661-667.
- Roskies, A. (2003). Are ethical judgments intrinsically motivational? Lessons from "acquired sociopathy". *Philosophical Psychology*, 16 (1), 51-66.
- Russell, D.C. (2009). *Practical Intelligence and the Virtues*. Oxford: Oxford University Press.
- Schein, C. y Gray, K. (2014). The prototype model of blame: Freeing moral cognition from linearity and little boxes. *Psychological Inquiry*, 25 (2), 236-240.
- Schein, C. y Gray, K. (2015). The Unifying Moral Dyad: Liberals and Conservatives Share the Same Harm-Based Template. *Personality and Social Psychology Bulletin*, 41(8), 1147-1163.
- Schnall, S., Roper, J. y Fessler, D. (2010). Elevation leads to altruistic behavior. *Psychological Science*, 21 (3), 315-320.
- Schroeder, T., Roskies, A. L. y Nichols, S. (2010). Moral Motivation. En J. M. Doris (Ed.), *The Moral Psychology Handbook*. Oxford: Oxford University Press.
- Schwitzgebel, E. y Ellis, J. (2017). Rationalization in Moral and Philosophical Thought. En J.F. Bonnefon y B. Tremoliere (Eds.), *Moral Inferences* (pp. D). Hove: Psychology Press.
- Searle, J. R. (1964). How to derive "ought" from "is". *Philosophical Review*, 73 (1), 43-58.
- Shafer-Landau, R. (2014). *The Fundamentals of Ethics*. Nueva York: Oxford University Press.
- Shaver, K.G. (1985). *The Attribution of Blame: Causality, Responsibility and Blameworthiness*. New York, NY: Springer.

- Shoemaker, D. (En prensa). Response-Dependent Responsibility; or, A Funny Thing Happened on the Way to Blame. *Philosophical Review*, 126 (4), 481-527.
- Silvers, J. A. y Haidt, J. (2008). Moral elevation can induce nursing. *Emotion*, 8 (2), 291–295.
- Sinnott-Armstrong, W. E. (2008b). Framing moral intuitions. En W. Sinnott-Armstrong (Ed.), *Moral Psychology: The Cognitive Science of Morality: Intuition and Diversity (Vol. 2)*. (pp. 47-82). Cambridge: MIT Press.
- Sinnott-Armstrong, W. E. (Ed.). (2008a). *Moral psychology: The evolution of morality: Adaptations and innateness (Vol. 1)*. Cambridge: MIT Press.
- Sinnott-Armstrong, W. E. (Ed.). (2008b). *Moral Psychology: The Cognitive Science of Morality: Intuition and Diversity (Vol. 2)*. Cambridge: MIT press.
- Sinnott-Armstrong, W. E. (Ed.). (2008c). *Moral psychology: The neuroscience of morality: Emotion, brain disorders, and development (Vol. 3)*. Cambridge: MIT Press.
- Sinnott-Armstrong, W. E. (Ed.). (2014). *Moral psychology: Free Will and Moral Responsibility (Vol. 4)*. Cambridge: MIT Press.
- Skulmowski, A., Bunge, A., Kaspar, K. y Pipa, G. (2014). Forced-choice decision-making in modified trolley dilemma situations: a virtual reality and eye tracking study. *Frontiers in Behavioral Neuroscience*, 8, 1-16.
- Slooman, S. A. y Lagnado, D. (2015). Causality in thought. *Annual Review of Psychology*, 66, 223-247.
- Slooman, S. A., Fernbach, P. M. y Ewing, S. (2009). Causal models: The representational infrastructure for moral judgment. *Psychology of Learning and Motivation*, 50, 1-26.
- Smetana, J. (1989). Toddler's social interactions in the context of moral and conventional transgressions in the home. *Developmental Psychology*, 25, 499–508.
- Smetana, J. y Braeges, J. (1990). The development of toddlers' moral and conventional judgements. *Merrill-Palmer Quarterly*, 36, 329-46.
- Smetana, J., Schlagman, N. y Walsh, P. (1993). Preschool Children's Judgments about Hypothetical and Actual Transgressions. *Child Development*, 64 (1), 202-214.
- Smith, A. M. (2005). Responsibility for attitudes: Activity and passivity in mental life. *Ethics*, 115 (2), 236-271.
- Sober, E. & Wilson, D.S. (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.
- Stich, S. (2007). Evolution, altruism and cognitive architecture: a critique of Sober and Wilson's argument for psychological altruism. *Biology & Philosophy*, 22(2), 267-281.
- Strawson, P.F. (1962/2013). Freedom and Resentment. En P. Russell y O. Deery (Eds.), *The philosophy of free will: Essential readings from the contemporary debates* (pp. 63-83). Oxford: Oxford University Press.
- Strohlinger, N. y Kumar, V. (Ed.) (2018). *The Moral Psychology of Disgust*. London: Rowman & Littlefield International.
- Teper, R., Inzlicht, M., & Page-Gould, E. (2011). Are we more moral than we think? Exploring the role of affect in moral behavior and moral forecasting. *Psychological Science*, 22 (4), 553-558.
- Thomson, J. J. (1976). Killing, Letting Die, and the Trolley Problem. *The Monist*, 59, 204-17.
- Thomson, J. J. (1985). The Trolley Problem. *The Yale Law Journal*, 94, 1395–1415.

- Trivers R. L. (2006) Reciprocal altruism: 30 years later. In: Kappeler P.M., van Schaik C.P. (Eds.) *Cooperation in Primates and Humans* (pp.67-83). Berlin: Springer.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46(1), 35-57.
- Turiel, E. (1983). *The Development of Social Knowledge: Morality and Convention*. Cambridge: Cambridge University Press.
- Turiel, E., Killen M. y Helwig, C. (1987). Morality: Its structure, functions, and vagaries. En J. Kagan y S. Lamb (Eds.), *The emergence of morality in young children* (pp.155-244). Chicago: University of Chicago Press.
- Uhlmann, E. L., Pizarro, D. A. y Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, 10 (1), 72-81.
- Vargas, M. (2007). Revisionism. En J. M. Fischer, R. Kane, D. Pereboom y M. Vargas (Eds.), *Four Views on Free Will* (pp. 126-165). Malden: Blackwell.
- Vargas, M. (2013a). *Building Better Beings: A theory of moral responsibility*. Oxford: Oxford University Press.
- Vargas, M. (2013b). Situationism and Moral Responsibility. En A. Clark, J. Kiverstein y T. Vierkant (Eds.), *Decomposing the Will. Free Will in Fragments* (pp. 325-343). Oxford: Oxford University Press.
- Wallace, R. J. (1994). *Responsibility and the Moral Sentiments*. Cambridge, Mass.: Harvard University Press.
- Wallace, R. J. (2000). *Responsibility and the moral sentiments*. Cambridge: Harvard University Press.
- Wason, P. C. y Evans, J. St. B. T. (1975). Dual processes in reasoning? *Cognition*, 3, 141-154.
- Watson, G. (1996). Two faces of responsibility. *Philosophical Topics*, 24 (2), 227-248.
- Watson, G. (2013). Responsibility and the Limits of Evil: Variations on a Strawsonian Theme. En P. Russell y O. Deery (Eds.), *The philosophy of free will: Essential readings from the contemporary debates* (pp. 84-113). Oxford: Oxford University Press.
- Wiegmann, A. y Waldmann, M. R. (2014). Transfer effects between moral dilemmas: A causal model theory. *Cognition*, 131 (1), 28-43.
- Woolfolk, R. L., Doris J. M. y Darley J. M. (2006). *Cognition*, 100 (2), 283-301.
- Young, L. y Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition*, 120 (2), 202-214.
- Zechmeister, J.S. y Romero, C. (2002). Victim and offender accounts of interpersonal conflict: autobiographical narratives of forgiveness and unforgiveness. *Journal of Personality and Social Psychology*, 82(4), 675-686.
- Zimmerman, M. J. (1988). *An Essay on Moral Responsibility*. Michigan: Roman & Littlefield.