RHV. An International Journal of Philosophy, 2022, No 20, 183-229 eISSN 0719-4242
DOI: https://doi.org/10.22370/rhv2022iss20pp183-229
Monographic Section

La broma secreta del alma de Kant[1]

The Secret Joke of Kant's Soul

Joshua Greene
Harvard University, Estados Unidos
jgreene@wjh.harvard.edu

[Traducción al español del texto: Greene, J. D. (2008). The secret joke of Kant's soul. En Sinnott-Armstrong, W. E. *Moral psychology: The neuroscience of morality: Emotion, brain disorders, and development*, 3, 35-79. MIT Press.]

> Dos cosas llenan el ánimo de admiración y respeto, siempre nuevos y crecientes, cuanto con más frecuencia y aplicación se ocupa de ellas la reflexión: el cielo estrellado que está sobre mí y la ley moral que está en mi interior. -— Immanuel Kant.
>
> Que semejante uso contranatural de las propias facultades sexuales (por tanto, abuso) viola el deber para consigo mismo, oponiéndose sin duda en sumo grado a la moralidad, es evidente para todo el mundo en cuanto piensa en él, y hasta tal punto suscita aversión a este pensamiento que incluso se tiene como inmoral mencionar un vicio semejante con su propio nombre (…) Ahora bien, no es tan fácil suministrar la prueba racional de que es inadmisible aquel uso contranatural de las propias facultades sexuales, e incluso simplemente el usarlas sin fin, en tanto que violación del deber para consigo mismo (y ciertamente, en lo que concierne al primero, en sumo grado). - El fundamento de la prueba consiste sin duda en que el hombre renuncia con ello (desdeñosamente) a su personalidad, al usarse únicamente como medio para satisfacer los impulsos animales. — Immanuel Kant.

[1] Traducción: E. Joaquín Suárez-Ruíz. Revisión: Pedro Pérez Zafrilla y Fernando Manzini. Esta traducción cuenta con la autorización tanto de Joshua Greene, como así también de Walter Sinnott-Armstrong (editor del volumen en el que se publicó originalmente el texto) y de la editorial original (MIT Press).

Received: 01/05/2022. Final version: 15/11/2022
eISSN 0719-4242 – © 2022 Instituto de Filosofía, Universidad de Valparaíso

> La broma de Kant–Kant quería probar, de una manera que ofendiera a "todo el mundo", que "todo el mundo" tenía razón: esa era la broma secreta de esta alma. Escribió contra los eruditos a favor del prejuicio popular, pero para los eruditos y no para el pueblo. — Friedrich Nietzsche.[2]

Existe evidencia sustancial y creciente que sugiere que gran parte de lo que hacemos, lo hacemos inconscientemente y por razones que nos resultan inaccesibles (Wilson, 2002). En un experimento, por ejemplo, se pidió a un grupo de personas que eligieran una de las varias medias presentadas en hilera. Cuando se les pidió que explicaran sus preferencias, las personas dieron respuestas razonables, refiriéndose a características relevantes de los elementos elegidos: su tejido superior, su transparencia, su elasticidad, etc. Pero sus elecciones no tenían nada que ver con tales características, ya que los objetos en exhibición eran, de hecho, idénticos. La gente simplemente tenía una preferencia por los elementos en el lado derecho de la pantalla (Nisbett y Wilson, 1977). Lo que ilustra este experimento (y hay muchos, muchos ejemplos de este tipo), es que las personas (1) toman decisiones por razones que desconocen y (2) inventan justificaciones razonables para sus elecciones, al mismo tiempo que desconocen sus motivaciones reales y sus racionalizaciones posteriores.

Jonathan Haidt, en su influyente artículo "El perro emocional y su cola racional: un enfoque intuicionista social del juicio moral" (Haidt 2001), aplica estas lecciones psicológicas al estudio del juicio moral. Haidt argumenta que, en su mayor parte, el razonamiento moral es *post-hoc*: decidimos lo que es correcto o incorrecto sobre la base de intuiciones impulsadas por las emociones y, si es necesario, inventamos razones para explicar y justificar nuestros juicios. Haidt admite que algunas personas, algunas veces, pueden llegar a conclusiones morales mediante el razonamiento, pero insiste en que esta no es la norma. Más importante aún para los propósitos de este ensayo, Haidt no distingue entre los diversos enfoques de la ética que resultan familiares a los filósofos morales: el consecuencialismo, la deontología, la ética de la virtud, etc. Más bien, su tesis radical está destinada, aunque solo sea implícitamente, a aplicarse por igual a los adherentes de todas las filosofías morales, aunque no necesariamente a los filósofos morales como grupo (Kuhn 1991).

[2] N. del T.: Para la traducción de los dos epígrafes de Kant se utilizó, respectivamente, las siguientes fuentes: Kant, I. (1994 [1788]). *Crítica de la razón práctica* (trad: E. Miñana Y Villasagra y Manuel García Morente). Salamanca: Sígueme; Kant, I. (2005 [1797]) (pp. 197). *La metafísica de las costumbres* (trad: Adela Cortina Orts y Jesús Conill Sancho), pp. 284. Madrid: Tecnos. El epígrafe de Nietzsche se tradujo del original en alemán: Kant's Witz. Kant wollte auf eine ‚alle Welt' vor den Kopf stossende Art beweisen, dass ‚alle Welt' Recht habe: – das war der heimliche Witz dieser Seele. Er schrieb gegen die Gelehrten zu Gunsten des Volks-Vorurtheils, aber für die Gelehrten und nicht für das Volk (Nietzsche, F. 1967. Sämtliche Werke. Kritische Studienausgabe in 15 Bänden (pp. 504). München: Deutscher Taschenbuch Verlag).

Jonathan Baron (Baron 1994), en contraste, hace una distinción psicológica entre juicios consecuencialistas y no consecuencialistas, argumentando que estos últimos son probablemente realizados sobre la base de heurísticas, reglas simples para la toma de decisiones. Sin embargo, el psicólogo no considera que la emoción sea esencial para estos juicios heurísticos.

En este ensayo me baso en las ideas de Haidt y Baron, al servicio de un poco de psicoanálisis filosófico. Argumentaré que los juicios deontológicos tienden a ser impulsados por respuestas emocionales y que la filosofía deontológica, en lugar de basarse en el razonamiento moral, consiste en gran medida un ejercicio de racionalización moral. Esto contrasta con el consecuencialismo, que, según sostendré, surge de procesos psicológicos bastante diferentes, los cuales son más "cognitivos" e involucran, con mayor probabilidad, un razonamiento moral genuino. Estas afirmaciones son estrictamente empíricas y las defenderé sobre la base de la evidencia disponible. Está de más decir que mi argumento será especulativo y no concluyente. Hecha la salvedad, argumentaré que si estas afirmaciones empíricas son ciertas, pueden tener consecuencias normativas: pondrían en duda la comprensión de la deontología como una escuela de pensamiento moral normativo.

1. Preliminares

1.1 Definiendo Deontología y Consecuencialismo

La deontología se define por su énfasis en las reglas morales, la mayoría de las veces articuladas en términos de derechos y deberes. El consecuencialismo, en contraste, supone que el valor moral de una acción es únicamente, de una u otra manera, una función de sus consecuencias. Los consecuencialistas sostienen que los responsables de la toma de decisiones morales siempre deben tratar de producir las mejores consecuencias generales para todos los involucrados, ya sea de manera directa o indirecta. Tanto los consecuencialistas como los deontólogos sostienen que las consecuencias son importantes. No obstante, mientras que los deontólogos creen que la moral requiere y nos permite hacer cosas que no producen las mejores consecuencias posibles, los consecuencialistas consideran que las consecuencias, en última instancia, son lo único que importa. Por ejemplo, un deontólogo podría decir que matar a una persona con el fin salvar a otras está mal, incluso si el hacerlo permite maximizar buenas consecuencias (Kagan 1997).

Lo anterior es una explicación estándar de qué son la deontología y el consecuencialismo y en qué se diferencian. A la luz de esta explicación, podría parecer que mi tesis es falsa por definición. La deontología es una moral basada en reglas, generalmente centrada en derechos y deberes. Un juicio deontológico, entonces, es un juicio realizado de acuerdo a ciertos tipos de reglas morales. De esto se deduce que un juicio moral elaborado sobre la base de una respuesta emocional, aunque a primera vista pueda parecerlo, simplemente no puede ser un juicio deontológico. El propio Kant se mostró inflexible sobre esto, al menos con respecto a su

propia forma de deontología. Afirmó manifiestamente que una acción realizada simplemente por simpatía y no por un reconocimiento del deber propio carece de valor moral (Kant 1959, Cap. 1; Korsgaard, 1996a Cap. 2).

La suposición detrás de esta objeción —que hasta donde yo sé nunca ha sido cuestionada previamente— es que el consecuencialismo y la deontología son, ante todo, filosofías morales. Se supone que los filósofos saben exactamente qué son la deontología y el consecuencialismo porque estos términos/conceptos fueron definidos por los filósofos. A pesar de ello, creo que es posible que los filósofos no sepan necesariamente qué son realmente el consecuencialismo y la deontología.

¿Cómo podría ser esto posible? La respuesta, según propondré, es que los términos «deontología» y «consecuencialismo» se refieren a clases naturales de tipo psicológico. Considero que las filosofías consecuencialista y deontológica no son tanto invenciones filosóficas como sí manifestaciones filosóficas de dos patrones psicológicos disociables, dos formas diferentes de pensamiento moral, que han sido parte del repertorio humano desde hace miles de años. Según este punto de vista, las filosofías morales de Kant, Mill y otros son solo las puntas explícitas de grandes icebergs psicológicos, en su mayoría implícitos. Si esto es correcto, entonces puede que los filósofos no sepan realmente con qué están lidiando cuando utilizan teorías morales consecuencialistas y deontológicas, y es posible que tengamos que hacer algo de ciencia para averiguarlo.

Comencemos con una analogía basada en un tema filosófico familiar: supongamos que en una determinada tierra tropical se refieren al agua con este símbolo: S. A su vez, en su Diccionario Sagrado se establece claramente que S es un líquido claro y potable (es decir, el diccionario define S en términos de su "intensión primaria" (Chalmers 1996)). Un día, una joven emprendedora viaja a la cima de una montaña cercana y se convierte en la primera de su gente en encontrar hielo. A través de un poco de experimentación, descubre que el hielo es una forma de agua y les cuenta con entusiasmo a los ancianos de la tribu sobre su descubrimiento. Al día siguiente, lleva a uno de los ancianos a la cima de la montaña, le entrega un poco de hielo y le dice: "¡Mira! ¡S!". En ese momento, el anciano exasperado le explica a la joven que S es un líquido, que la sustancia dura que tiene en la mano claramente no es un líquido y que no le gusta que le hagan perder el tiempo.

En un sentido estricto, el anciano tiene razón. El Diccionario Sagrado es la autoridad sobre lo que significan los símbolos locales, y establece muy claramente que S se refiere a un líquido claro y bebible. Pero al anciano le falta el panorama general. Lo que está olvidando, o tal vez nunca entendió, es que muchas cosas en el mundo tienen estructuras subyacentes ("esencias", si se prefiere), las cuales son responsables de hacer que las cosas aparezcan y se comporten como lo hacen, esto es, de darles sus propiedades funcionales. De hecho, es debido a que las cosas tienen estructuras subyacentes que es posible referirse a algo, incluso inventar una definición de ello, sin comprender realmente qué es (Kripke 1980; Putnam 1975). Por supuesto, una comunidad lingüística podría insistir en que su definición es correcta. Nadie

debe impedirles utilizar sus símbolos como les plazca. Pero, al hacer esto, corren el riesgo de perder el panorama general, de negarse a sí mismos una comprensión más profunda de lo que sucede a su alrededor, o incluso dentro de ellos.

Debido a que estoy interesado en explorar la posibilidad de que la deontología y el consecuencialismo sean clases psicológicas naturales, dejaré de lado sus definiciones filosóficas convencionales y me enfocaré en sus roles funcionales relevantes. Como se señaló anteriormente, los consecuencialistas y los deontólogos tienen algunos desacuerdos prácticos característicos. Por ejemplo, los consecuencialistas suelen decir que matar a una persona para salvar a otras puede ser lo correcto, dependiendo de la situación. Los deontólogos, por el contrario, suelen decir que está mal matar a una persona en beneficio de otros, es decir, que los "fines no justifican los medios". Dado que los consecuencialistas y los deontólogos tienen este tipo de desacuerdos prácticos, podemos usarlos para definir funcionalmente los juicios consecuencialistas y los deontológicos. A los fines de esta discusión, afirmaremos que los juicios consecuencialistas son juicios a favor de conclusiones característicamente consecuencialistas (por ejemplo, "Es mejor salvar más vidas") y que los juicios deontológicos son juicios a favor de conclusiones característicamente deontológicas (por ejemplo, "Es incorrecto a pesar de los beneficios"). Mi uso de "característicamente" es obviamente poco preciso aquí, pero confío en que aquellos familiarizados con los debates éticos contemporáneos sabrán lo que quiero decir. Tenga en cuenta que el tipo de juicio realizado es en gran medida independiente de quién lo está haciendo. Un deontólogo militante puede emitir un juicio "característicamente consecuencialista", como cuando Judith Jarvis Thomson afirma que está bien desviar un tranvía fuera de control que amenaza con matar a cinco personas, con el fin de que mate a solo una persona (Thomson 1986). Este es un juicio "característicamente consecuencialista" porque se justifica fácilmente en términos de los principios consecuencialistas más básicos, mientras que los deontólogos deben realizar un montón de sofisticados argumentos filosóficos con el fin de defender esta posición. Del mismo modo, considere el juicio de que es incorrecto salvar a cinco personas que necesitan trasplantes de órganos quitándole los órganos a un donante que no dio su consentimiento (Thompson 1986). Este juicio es "característicamente deontológico", no porque muchos consecuencialistas militantes no estén de acuerdo, sino porque tienen que dar muchas explicaciones adicionales para justificar su acuerdo.

Al definir el "consecuencialismo" y la "deontología" en términos de sus juicios característicos, le damos una oportunidad a nuestra hipótesis empírica. Si resulta que los juicios característicamente deontológicos son impulsados por la emoción (una posibilidad empírica), entonces se plantea la posibilidad de que la filosofía deontológica también sea impulsada por la emoción (una posibilidad empírica adicional). En otras palabras, lo que encontramos cuando exploramos las causas psicológicas de los juicios característicamente deontológicos podría sugerir que lo que la filosofía moral deontológica realmente es, lo que es en esencia, es un intento de producir justificaciones racionales para juicios morales impulsados emocionalmente y no un intento de llegar a conclusiones morales sobre la base del razonamiento moral.

Sin embargo, el punto por ahora es simplemente señalar el problema terminológico. Cuando me refiero a algo como un "juicio deontológico", estoy diciendo que es un juicio característicamente deontológico y no que el juicio en cuestión cumpla necesariamente con los criterios que los filósofos impondrían al considerar ese juicio como deontológico. Al final, sin embargo, argumentaré que tales juicios se entienden mejor como genuinamente deontológicos porque son producidos por una psicología subyacente, que es la esencia oculta de la filosofía deontológica.

1.2 Definiendo "cognición" y emoción

En lo que sigue, argumentaré que mientras el juicio deontológico tiende a ser impulsado por la emoción, el juicio consecuencialista tiende a ser impulsado por procesos "cognitivos". ¿Qué entendemos por "emoción" y "cognición" y en qué se diferencian?

A veces, "cognición" se refiere al procesamiento de información en general, como en "ciencia cognitiva", pero a menudo "cognición" se usa en un sentido más estrecho que contrasta con "emoción", a pesar del hecho de que las emociones involucran procesamiento de información. No conozco una buena definición de "cognición" en este sentido más restrictivo, a pesar de su extendido uso. En otro trabajo, mis colaboradores y yo ofrecimos una definición tentativa propia (Greene et al. 2004), una que se basa en las diferencias entre los requisitos del procesamiento de información del comportamiento estereotipado frente al comportamiento flexible.

La idea general es que las representaciones "cognitivas" son inherentemente neutrales, es decir, no desencadenan automáticamente respuestas o disposiciones de comportamiento particulares. Por otro lado, las representaciones "emocionales" son aquellas que sí tienen tales efectos automáticos y, a su vez, tienen una valencia comportamental. El comportamiento altamente flexible requiere representaciones "cognitivas" que pueden mezclarse y combinarse fácilmente según lo que la situación demande, y esto sin la necesidad de tironear al agente en dieciséis direcciones de comportamiento diferentes a la vez. Por ejemplo, a veces es necesario evitar los automóviles y otras veces resulta preciso acercarse a ellos. Resulta útil, entonces, el poder representar AUTO de manera neutral o "cognitiva", es decir, de un modo que no presupone automáticamente una respuesta conductual en particular. El comportamiento estereotipado, en contraste, no requiere este tipo de flexibilidad y, por lo tanto, no requiere representaciones "cognitivas", al menos no en la misma medida. Para dejarlo claro usaré las comillas cuando use el sentido más restrictivo, aquí definido, de "cognitivo" y las quitaré cuando lo use para referirme al procesamiento de información en general.

Si bien todo el cerebro está dedicado a la cognición, los procesos "cognitivos" son especialmente importantes para el razonamiento, la planificación, la manipulación de información en la memoria de trabajo, el control de los impulsos y las "funciones ejecutivas superiores" en general. Además, estas funciones tienden a estar asociadas con ciertas partes del cerebro,

principalmente con las superficies dorsolaterales de la corteza prefrontal y los lóbulos parietales (Koechlin et al. 2003; Miller y Cohen 2001; Ramnani y Owen 2004). La emoción, en contraste, tiende a asociarse con otras partes del cerebro, como la amígdala y las superficies mediales de los lóbulos frontal y parietal (Adolphs 2002; Maddock 1999; Phan et al. 2002). Y mientras que el término "emoción" puede referirse a estados estables como los estados de ánimo, aquí nos ocuparemos principalmente de las emociones provocadas por procesos que, además de tener una valencia particular, son rápidos y automáticos, aunque no necesariamente conscientes.

Aquí nos ocuparemos, entonces, de dos tipos diferentes de juicio moral (deontológico y consecuencialista) y dos tipos diferentes de proceso psicológico ("cognitivo" y emocional). Cruzándolos obtenemos cuatro posibilidades empíricas básicas. Primero, podría ocurrir que ambos tipos de juicio moral sean generalmente "cognitivos", como sugieren las teorías de Kohlberg (Kohlberg 1971)[3]. En el otro extremo, podría suceder que ambos tipos de juicio moral sean principalmente emocionales, como sugiere el enfoque de Haidt (Haidt 2001). Luego está el estereotipo histórico, según el cual el consecuencialismo es más emocional (que surge de la tradición "sensibilista" de David Hume [Hume 1978] y Adam Smith [Smith 1976]) y la deontología es más "cognitiva" (la cual abarca la tradición "racionalista" kantiana [Kant 1959]). Finalmente, hay un enfoque, a favor del cual argumentaré, que sostiene que la deontología está más motivada por las emociones y el consecuencialismo es más "cognitivo". Sin embargo, adelanto que no creo que ninguno de los enfoques sea estrictamente emocional o "cognitivo" (o incluso que hay una clara distinción entre "cognición" y emoción). Más específicamente, simpatizo con la afirmación de Hume de que todo juicio moral (incluido el juicio consecuencialista) debe tener algún componente emocional (Hume 1978). No obstante, sospecho que el tipo de emoción que es esencial para el consecuencialismo es fundamentalmente diferente del tipo que es esencial para la deontología: mientras que para el primero la emoción funciona como un elemento habitual y conocido, para la segunda funciona como una alarma. Volveremos a este problema más adelante.

2. Evidencia científica

2.1 Evidencia de Neuroimagen

En las últimas décadas, los filósofos han ideado una serie de dilemas morales hipotéticos que capturan la tensión entre los puntos de vista consecuencialista y deontológico. Un puñado bien conocido de estos dilemas da lugar a lo que se conoce como el "problema del tranvía" [*trolley problem*] (Foot 1978; Thomson 1986), que comenzó con el siguiente:

[3] Kohlberg era partidario de la deontología y probablemente diría que esta es más "cognitiva" que el consecuencialismo.

Un tranvía fuera de control se dirige a cinco personas que serán asesinadas si continúa en su curso actual. La única forma de salvarlas es presionando un interruptor que desviará el tranvía hacia una vía lateral donde arrollará y matará a una persona en lugar de a cinco. ¿Está bien desviar el tranvía para salvar a cinco personas a costa de una? El consenso entre los filósofos (Fischer y Ravizza 1992), así como entre las personas que han sido parte de experimentos relacionados (Petrinovich y O'Neill 1996; Petrinovich et al. 1993), es que, en este caso, resulta moralmente aceptable salvar cinco vidas a expensas de una.

A continuación consideremos el dilema del puente peatonal [*footbridge dilemma*] (Thomson 1986): como en el caso anterior, un tranvía fuera de control amenaza con matar a cinco personas, pero esta vez usted está de pie junto a un desconocido de complexión física grande en un puente peatonal que cruza las vías, el cual se encuentra entre el carro que se aproxima y las cinco personas. La única forma de salvar a las cinco personas es empujando al extraño del puente, de modo tal que caiga en las vías de abajo. Como resultado, dicha persona morirá, pero su cuerpo evitará que el tranvía alcance a los demás. ¿Está bien salvar a las cinco personas empujando a este extraño a su muerte? Aquí el consenso es que no está bien salvar cinco vidas a costa de una (Fischer y Ravizza 1992; Greene et al. 2004; Greene et al. 2001; Petrinovich y O'Neill 1996; Petrinovich et al. 1993).

Las personas presentan una respuesta característicamente consecuencialista en el caso del tranvía y una respuesta característicamente deontológica en el caso del puente. ¿Por qué? Los filósofos generalmente han ofrecido una variedad de explicaciones normativas. Es decir, han asumido que nuestras respuestas a estos casos son correctas, o al menos razonables, y han buscado principios que justifiquen el tratarlos de manera diferente (Fischer y Ravizza 1992). Por ejemplo, uno podría suponer, siguiendo a Kant (1959) y a Tomás de Aquino (1988), que es incorrecto dañar a alguien como un medio para ayudar a alguien más. En el caso del puente, la acción propuesta implica literalmente utilizar a la persona en el puente como un obstáculo; en el caso del tranvía, por su parte, la víctima es dañada simplemente como un efecto secundario (si esa persona de pronto desapareciera mágicamente, estaríamos encantados). En respuesta a esta propuesta, Thomson ideó el caso del bucle [*loop case*] (Thomson 1986). Aquí, la situación es similar a la del dilema del tranvía, pero esta vez la persona se encuentra en un tramo que se sale de la vía principal y luego se reincorpora a ella un poco antes de las cinco personas. En este caso, si la persona no estuviera en la vía lateral, el carro volvería a la vía principal y pasaría por encima de las cinco personas. El consenso en este caso es que resulta moralmente aceptable girar el tranvía, a pesar del hecho de que aquí, como en el caso de la pasarela, una persona estaría siendo utilizada como un medio.

Ha habido muchos intentos normativos de este tipo con el fin de resolver el problema del tranvía, pero ninguno de ellos ha sido tremendamente exitoso (Fischer y Ravizza 1992). Mis colaboradores y yo hemos propuesto una solución parcial y puramente descriptiva al problema y hemos recopilado algunas pruebas científicas a su favor. Planteamos la hipótesis de que la idea de empujar a alguien a su muerte de una manera "cercana y personal" (como sucede en el dilema del puente) es más emocionalmente significativo que la idea de provocar consecuen-

cias similares de una manera más impersonal (por ejemplo, al presionar un interruptor, como sucede en el dilema del tranvía). Propusimos que esta diferencia en la respuesta emocional explica por qué las personas responden de manera tan diferente a estos dos casos. Es decir, las personas tienden hacia el consecuencialismo en el caso de que la respuesta emocional sea baja y a la deontología en el caso en que la respuesta emocional sea alta.

La razón para distinguir entre formas de daño personal e impersonal es en gran medida evolutiva. La violencia "cercana y personal" existe desde hace mucho tiempo y, de hecho, se remonta a nuestro linaje de primates (Wrangham y Peterson 1996). Dado que la violencia personal es evolutivamente antigua, existe previamente a nuestras capacidades humanas recientemente desarrolladas para el razonamiento abstracto complejo. Por tanto, no debería sorprendernos si tenemos respuestas innatas a la violencia personal que son poderosas y, al mismo tiempo, bastante primitivas. Es decir, podríamos esperar que los humanos tengan respuestas emocionales negativas a ciertas formas básicas de violencia interpersonal, ya que estas respuestas evolucionaron como un medio para regular el comportamiento de aquellas criaturas que son capaces de dañarse intencionalmente entre sí, pero cuya supervivencia depende de la cooperación y de la restricción individual (Sober y Wilson 1998; Trivers 1971). Por el contrario, cuando un daño es impersonal, no lograría activar esta respuesta emocional de tipo alarma, habilitando que las personas respondan de una manera más "cognitiva", tal vez empleando un análisis de costo-beneficio. Como dijo una vez Josef Stalin: "Una sola muerte es una tragedia; un millón de muertes es una estadística". Su comentario sugiere que cuando las acciones dañinas son lo suficientemente impersonales, no son capaces de presionar nuestros botones emocionales, incluso a pesar de su seriedad, y como resultado pensamos en ellas de una manera actuarial, más indiferente.

Esta hipótesis sugiere algunas predicciones sólidas con respecto a lo que deberíamos ver en el cerebro de las personas mientras responden a dilemas relacionados con daños personales e impersonales (de aquí en adelante dilemas morales "personales" e "impersonales"). Mientras que la contemplación de dilemas morales personales, como el caso del puente, debería producir una mayor actividad neuronal en las regiones cerebrales asociadas con la respuesta emocional y la cognición social, la contemplación de dilemas morales impersonales, como el caso del tranvía, debería producir una actividad relativamente mayor en las regiones cerebrales asociadas con la "cognición superior"[4]. Esto es exactamente lo que sucedió (Greene et al. 2004; Greene et al. 2001). La contemplación de los dilemas morales personales produjo una

[4] Determinar qué hace que un dilema moral sea "personal" y "como el caso del puente peatonal" frente a "impersonal" y "como el caso del tranvía" no es un asunto sencillo. De hecho, en muchos sentidos reintroduce las complejidades asociadas con los intentos tradicionales de resolver el problema del tranvía. Sin embargo, a los fines de esta discusión, prefiero suponer la distinción personal-impersonal como intuitiva, de acuerdo con la descripción evolutiva desarrollada anteriormente. Vale resaltar, no obstante, que a los efectos de diseñar el experimento de imágenes cerebrales analizado más adelante, mis colaboradores y yo desarrollamos un conjunto de criterios más rígidos para distinguir las violaciones morales personales de las impersonales (Greene et al. 2001). Creo que estos criterios ya no son adecuados. Mejorarlos es un objetivo de la investigación en curso.

actividad relativamente mayor en tres áreas relacionadas con las emociones: la corteza cingulada posterior, la corteza prefrontal medial y la amígdala. Este efecto también se observó en el surco temporal superior, una región asociada con varios tipos de cognición social en humanos y otros primates (Allison et al. 2000; Saxe et al. 2004). Al mismo tiempo, la contemplación de dilemas morales impersonales produjo una actividad neuronal relativamente mayor en dos áreas cerebrales clásicamente "cognitivas": la corteza prefrontal dorsolateral y el lóbulo parietal inferior.

Esta hipótesis también sugiere una predicción sobre los tiempos de reacción de las personas. De acuerdo con la visión que he esbozado, las personas tienden a tener respuestas emocionales a las violaciones morales personales que las inclinan a juzgar en contra de la realización de esas acciones. Esto significa que si alguien juzga que una violación moral personal es apropiada (por ejemplo, alguien que dice que está bien empujar al hombre para que caiga del puente), probablemente tendrá que anular una respuesta emocional para poder hacerlo. Este proceso de anulación llevará tiempo y, por lo tanto, podríamos esperar que las respuestas "sí" tarden más que las "no" cuando se decide en el contexto de dilemas morales personales como el caso del puente. Al mismo tiempo, no tenemos ninguna razón para predecir una diferencia en el tiempo de reacción entre las respuestas de "sí" y "no" en relación con los dilemas morales impersonales como en el caso del tranvía, dado que, según este modelo, no hay una respuesta emocional (o incluso mucho menos de una) para anular en tales casos. Aquí también se mantuvo la predicción. Las pruebas en las que un sujeto juzgaba a favor de las violaciones morales personales tomaban significativamente más tiempo que las pruebas en las que juzgaba en contra de ellas, pero no se observó un efecto de tiempo de reacción comparable frente a las violaciones morales impersonales (Greene et al. 2004; Greene et al. al. 2001).

Hay otros resultados que también apoyan este modelo. A continuación, subdividimos los dilemas morales personales en dos categorías en función de su dificultad (es decir, basado en el tiempo de reacción). Considera el siguiente dilema moral (el dilema del bebé que llora): es tiempo de guerra, tú y algunos de tus vecinos se esconden de los soldados enemigos en un sótano. Tu bebé comienza a llorar y le cubres la boca para bloquear el sonido. Si retiras la mano, tu bebé llorará con fuerza, los soldados lo escucharán, los encontrarán y matarán a todos los que encuentren, incluidos a tu bebé y a ti. Si no quitas la mano, tu bebé se ahogará hasta morir. ¿Está bien asfixiar a tu bebé hasta matarlo para salvar tu vida y la de los demás aldeanos?

Se trata de una pregunta muy difícil. Las personas dan diferentes respuestas y casi todos se toman un tiempo relativamente largo. Esto contrasta con otros dilemas morales personales, como el dilema del infanticidio, en el que una adolescente debe decidir si matará o no a su recién nacido no deseado. En respuesta a este caso, las personas (al menos las que evaluamos) expresan rápida y unánimemente que esta acción es incorrecta.

¿Qué está pasando en estos dos casos? Mis colegas y yo planteamos la siguiente hipótesis. En ambos hay una respuesta emocional prevalente y negativa a la violación personal en cuestión, es decir, matar al bebé. En el caso del bebé que llora, sin embargo, hay un análisis

costo-beneficio que termina por favorecer fuertemente la opción de asfixiar al bebé. Después de todo, el bebé va a morir en cualquiera de las dos situaciones posibles y, por lo tanto, la persona no tiene nada que perder (en términos consecuencialistas) pero sí mucho que ganar asfixiándolo, por terrible que sea. En algunas personas, la respuesta emocional es la que domina y terminan diciendo "no". En otras, el análisis "cognitivo" costo-beneficio mencionado más arriba gana y dicen "sí".

¿Qué predice este modelo respecto de lo que podría verse en el cerebro de los participantes cuando se comparan casos como el del bebé llorando y el del infanticidio? En primer lugar, este modelo supone que casos como el del bebé que llora involucran un mayor nivel de «conflicto de respuesta», es decir, un conflicto entre representaciones en competencia que darían lugar a una respuesta conductual. En consecuencia, debemos esperar que los dilemas morales difíciles (como el del bebé que llora), produzcan un aumento de la actividad en una región del cerebro asociada con el conflicto de respuesta: la corteza cingulada anterior (Botvinick et al. 2001). En segundo lugar, según nuestro modelo, la diferencia crucial entre casos como los del bebé que llora y los del infanticidio es que los primeros evocan respuestas "cognitivas" fuertes que pueden competir efectivamente con una respuesta emocional prevalente. Por lo tanto, cuando comparamos casos como el del bebé que llora con el del infanticidio, deberíamos esperar ver una mayor actividad en áreas del cerebro "cognitivas" clásicas, a pesar del hecho de que dilemas difíciles (como el del bebé que llora) son dilemas morales personales que previamente se asociaron con una respuesta emocional (Greene et al. 2001).

Estas dos predicciones se han mantenido (Greene et al. 2004). La comparación de los dilemas morales personales de alto tiempo de reacción, como el del bebé que llora, con los dilemas morales personales de bajo tiempo de reacción, como el del infanticidio, reveló una mayor actividad en la corteza cingulada anterior (conflicto) así como también en la corteza prefrontal dorsolateral anterior y en los lóbulos parietales inferiores, ambas regiones cerebrales clásicamente "cognitivas".

Casos como el del bebé que llora son especialmente interesantes porque nos permiten comparar de manera directa la actividad neuronal asociada con las respuestas característicamente consecuencialistas y deontológicas. Según nuestro modelo, cuando las personas dicen "sí" en tales casos (la respuesta consecuencialista) se debe a que el análisis "cognitivo" de costo-beneficio ha dominado con éxito por sobre la respuesta emocional prevalente que impulsa a las personas a decir "no" (la respuesta deontológica). Si esto es correcto, entonces, deberíamos esperar ver un aumento de la actividad en las regiones cerebrales "cognitivas" previamente identificadas (corteza prefrontal dorsolateral y corteza parietal), en las pruebas en las que las personas dicen "sí" a casos como el del bebé que llora. Esto es exactamente lo que encontramos. En otras palabras, las personas exhiben más actividad "cognitiva" cuando dan la respuesta consecuencialista[5].

[5] Vale la pena señalar que ninguna región del cerebro, incluidas las implicadas en la emoción, evidenció el efecto contrario. En primer lugar, no está claro que pueda esperarse ver tal resultado, ya que la hipótesis es que todos los

En resumen, los juicios morales parecen ser producto de al menos dos tipos diferentes de procesos psicológicos. En primer lugar, tanto las imágenes cerebrales como los datos del tiempo de reacción sugieren que existen respuestas emocionales negativas prevalentes que llevan a las personas a desaprobar las propuestas que involucran acciones dañinas personales, a saber, casos como los dilemas del puente y del bebé que llora. Estas respuestas son características de la deontología, pero no del consecuencialismo. En segundo lugar, los resultados de las imágenes cerebrales adicionales sugieren que los procesos psicológicos "cognitivos" pueden competir con los procesos emocionales antes mencionados, lo que lleva a las personas a aprobar las violaciones morales dañinas a nivel personal, principalmente cuando existe una fuerte razón consecuencialista para hacerlo, como en el caso del bebé que llora. Cuando las personas hacen juicios característicamente consecuencialistas, las partes del cerebro que exhiben una mayor actividad son aquellas que están más estrechamente asociadas con funciones cognitivas superiores, como el control ejecutivo (Koechlin et al. 2003; Miller y Cohen 2001), la planificación compleja (Koechlin et al. al. 1999), el razonamiento deductivo e inductivo (Goel y Dolan 2004), la visión a largo plazo en la toma de decisiones económicas (McClure et al. 2004), etc. Además, en comparación con otros primates, estas regiones del cerebro se encuentran entre aquellas que se expandieron más dramáticamente en humanos (Allman et al. 2002).

2.2 La emoción y el sentido de la obligación moral

En su clásico artículo, "Hambruna, opulencia y moralidad", Peter Singer (1972) sostiene que nosotros, en el mundo rico, tenemos la obligación de hacer mucho más de lo que hacemos para mejorar las vidas de las personas necesitadas. Sostiene que, si podemos prevenir algo muy malo sin que eso signifique incurrir en un costo moral comparable, entonces deberíamos hacerlo. Por ejemplo, si uno nota que un niño pequeño se ahoga en un estanque poco profundo, uno está moralmente obligado a meterse y salvar a ese niño, incluso si eso significa ensuciarse la ropa. Como señala Singer, este principio aparentemente inocuo tiene implicaciones radicales, ya que conlleva que todos aquellos que gastan dinero en lujos innecesarios deban renunciar a ellos para así destinar el dinero a salvar y/o mejorar las vidas de las personas desfavorecidas. ¿Por qué, pregunta Singer, tenemos una obligación estricta de salvar a un niño cercano que se está ahogando, pero no una obligación comparable de salvar a los niños enfermos y hambrientos a través de donaciones caritativas a organizaciones como Oxfam?

Me vienen a la mente muchas explicaciones normativas, pero ninguna me parece demasiado convincente. ¿Se nos permite ignorar la difícil situación de los niños que están lejos de nosotros porque son ciudadanos de naciones extranjeras? En caso afirmativo, ¿sería aceptable dejar que el niño se ahogara si uno se encuentra con él mientras se está de viaje en el extran-

individuos experimentan una respuesta emocional intuitiva y solo algunos la anulan. En segundo lugar, es difícil sacar conclusiones de los resultados negativos de las neuroimágenes porque las técnicas actuales, que rastrean los cambios en el flujo sanguíneo, son instrumentos relativamente rudimentarios cuando se busca detectar patrones en la función neural.

jero? ¿O en aguas internacionales? ¿Y qué pasa con los pobres a nivel local? Este argumento no nos exime de nuestras obligaciones para con ellos. ¿Se debe acaso a una responsabilidad difusa, por el hecho de que muchos están en condiciones de ayudar a un niño hambriento en el extranjero pero solo usted está en condiciones de ayudar a este hipotético niño que se está ahogando? ¿Y si hubiera mucha gente parada alrededor del estanque sin hacer nada? ¿Eso lo habilitaría a no hacer nada también? ¿Se debe a que la ayuda internacional es en última instancia ineficaz y solo sirve para enriquecer a políticos corruptos o crear más gente pobre? En ese caso, nuestra obligación simplemente se trasladaría a esfuerzos de auxilio más sofisticados que incorporen reformas políticas, desarrollo económico, educación en planificación familiar, etc. ¿Están todos los esfuerzos de auxilio condenados a la ineficacia? Esa es una afirmación empírica audaz que nadie puede hacer con gran confianza de manera honesta.

Nos encontramos aquí en una situación similar a la que enfrentamos con el problema del tranvía. Tenemos la fuerte intuición de que dos dilemas morales son muy diferentes y, sin embargo, nos cuesta explicar cuál es esa importante diferencia (Kagan 1989; Unger 1996). Resulta que la misma teoría psicológica que da sentido al problema del tranvía puede dar sentido al problema desarrollado por Singer. Es preciso tener en cuenta que en el caso del niño que se ahoga la interacción es "cercana y personal", se trata de un tipo de situación con la que podrían haberse encontrado nuestros antepasados humanos y primates. A su vez, es importante considerar que el caso de la donación no es "cercano y personal", ni tampoco es un tipo de situación con la cual nuestros antepasados podrían haberse encontrado. No existía la posibilidad de que nuestros ancestros tuviesen la posibilidad de salvarles la vida a extraños anónimos mediante modestos sacrificios materiales. A la luz de esto, la teoría psicológica anteriormente presentada sugiere que es probable que encontremos más apremiante la obligación de salvar al niño que se está ahogando simplemente porque se trata de un caso "cercano y personal", el cual presiona nuestros botones emocionales de una manera en la que no lo hace el caso más impersonal de la donación (Greene 2003). Da la casualidad de que estos dos casos se encontraban entre los evaluados en el estudio de imágenes cerebrales descrito anteriormente, con una variación en el caso del niño ahogado incluido en la condición personal y con el caso de donación incluido en la condición impersonal (Greene et al. 2004; Greene et al. 2001).

Son pocos los que aceptan la conclusión consecuencialista de Singer. Más bien, las personas tienden a creer, de una manera característicamente deontológica, que el gastar su dinero en lujos para sí mismas está dentro de sus derechos morales, a pesar de que ese dinero podría usarse para mejorar considerablemente la vida de otros. Esto es exactamente lo que uno esperaría (1) si el sentido deontológico de obligación fuese impulsado principalmente por la emoción y (2) si se tratara de obligaciones de auxilio, es decir, si las emociones solo estuvieran lo suficientemente comprometidas cuando se encuentran a aquellos a quienes podríamos deberles algo (o que podría comprenderse de tal modo) de una forma personal.

2.3 La emoción y la atracción de víctimas identificables

Uno de los aspectos importantes para que una persona sea "cercana y personal" es que siempre debe tratarse, en cierto sentido, de un individuo identificable y determinado, y no de un individuo estadístico (Greene y Haidt 2002; Greene et al. 2001). Mientras que el niño que se está ahogando, por ejemplo, se presenta como una persona en particular, los niños a los que usted podría ayudar mediante donaciones a Oxfam son anónimos y, hasta donde usted sabe, indeterminados[6]. Muchos han observado una tendencia a responder con mayor urgencia a víctimas identificables, en comparación con víctimas "estadísticas" e indeterminadas (Schelling 1968). Esto se conoce como el "efecto de la víctima identificable".

Podría recordarse, por ejemplo, el caso de Jessica McClure, también conocida como "Baby Jessica", que en 1987 quedó atrapada en un pozo en Texas. Se enviaron más de $ 700,000 a su familia en apoyo al esfuerzo de rescate (Small y Loewenstein 2003; Variety 1989). Como señalan Small y Loewenstein, si esa cantidad de dinero se gastara en atención médica preventiva, podría haberse utilizado para salvar la vida de muchos niños. Esta observación plantea una cuestión normativa que es esencialmente la misma que la de Singer. ¿Tenemos una obligación mayor de ayudar a personas como Baby Jessica que de ayudar a un gran número de personas que podrían salvarse por menos? Si todo lo demás se mantiene igual, mientras que un consecuencialista diría "no", la mayoría de la gente aparentemente diría "sí". Además, la mayoría de las personas, si se las presiona para que expliquen su posición, es probable que lo hagan en términos deontológicos. Es decir, probablemente dirían que, por un lado, tenemos el deber de ayudar a alguien como Baby Jessica incluso si hacerlo implica un gran esfuerzo y gasto y que, por otro lado, no tenemos un deber comparable con los innumerables otros que podrían recibir ayuda utilizando los mismos recursos.

La misma teoría "cercana y personal" del compromiso emocional puede explicar este patrón de juicio. Otros han propuesto lo que equivale a la misma hipótesis y otros incluso han reunido evidencia independiente en apoyo a ella. En el artículo clave de Thomas Schelling sobre este tema, el economista observa que la muerte de una persona en particular invoca "ansiedad y sentimiento, culpa y asombro, responsabilidad y religión, [pero]... la mayor parte de esta maravilla desaparece cuando nos ocupamos de la muerte estadística" (Schelling 1968; Small y Loewenstein 2003). Inspirándose en la observación de Schelling, Small y Loewenstein realizaron dos experimentos destinados a probar la hipótesis de que "las víctimas identificables estimulan una respuesta emocional más poderosa que las víctimas estadísticas".

El aspecto crucial de su propuesta fue diseñar sus experimentos de tal manera que sus resultados pudieran contrarrestar todas las explicaciones normativas del efecto de la víctima identificable, es decir, explicaciones que buscan atribuir razones normativamente respetables a los responsables de la toma de decisiones cuando favorecen a víctimas identificables. Esto

[6] Por supuesto, algunas organizaciones de auxilio deliberadamente emparejan donantes individuales con beneficiarios individuales, con el fin de que la experiencia sea más personal.

es difícil porque el proceso de identificación de una víctima inevitablemente proporciona información sobre esa víctima (nombre, edad, sexo, apariencia, etc.), la cual podría servir como base racional para favorecer a esa persona. Con el fin de evitar esto, buscaron documentar una forma más débil del efecto de la víctima identificable, que podría llamarse el "efecto de la víctima determinada". Examinaron la disposición de las personas a beneficiar a individuos determinados frente a individuos indeterminados, en condiciones en las que toda la información significativa sobre las víctimas se mantiene constante.

Su primer experimento consistió en lo siguiente. A diez sujetos se les dio una "dotación" de diez dólares. Sacaban cartas al azar donde algunas decían "CONSERVAR", las cuales les permitían retener sus dotaciones, y otras decían "PERDER", las cuales ocasionaban que posteriormente se les quite sus dotaciones, convirtiéndolos así en "víctimas". Cada uno de los sujetos no víctimas fue emparejado de forma anónima con una de las víctimas, como resultado de haber sacado el número correspondiente a esa víctima. A los sujetos no víctimas se les permitió proporcionar una parte de sus donaciones a sus respectivas víctimas y cada uno podía elegir cuánto dar. Sin embargo, y esta es la manipulación crucial, mientras que algunos sujetos no víctimas sacaron el número de la víctima antes de decidir cuánto dar, otros sacaron el número de la víctima después de decidir, sabiendo de antemano que lo harían más tarde. En otras palabras, algunos sujetos tuvieron que responder a la pregunta "¿Cuánto le quiero dar a la persona #4?" (la víctima determinada) y otros sujetos tenían que responder a la pregunta "¿Cuánto quiero darle a la persona cuyo número sacaré?" (la víctima indeterminada). En ningún momento los sujetos no víctimas supieron quién recibiría su dinero. Los resultados: la donación media del grupo que donaba a víctimas determinadas fue un sesenta por ciento más alta que la del grupo que donaba a víctimas no determinadas. La mediana para el grupo de víctimas determinadas fue más del doble.

Vale la pena enfatizar lo absurdo de este patrón de comportamiento. No existe una base racional para darle más dinero a la "persona n°4 determinada al azar" que a la "persona n°? que se determinará aleatoriamente" y, sin embargo, eso es lo que hicieron estas personas[7] (es importante tener en cuenta que el experimento está diseñado para que ninguno de los participantes sepa quién eligió qué). ¿Por qué la gente haría esto? Aquí también la respuesta involucra

[7] En primer lugar, cuando digo que este comportamiento no se puede defender racionalmente no quiero decir que sea lógica o metafísicamente imposible que una persona racional se comporte de esta manera. Alguien podría, por ejemplo, tener una preferencia básica por ayudar solo a determinadas víctimas. Estoy asumiendo, sin embargo, que ninguno de los sujetos en este experimento tiene preferencias tan extrañas y que, por lo tanto, su comportamiento es irracional. En segundo lugar, no estoy afirmando que la tendencia psicológica general que produce este comportamiento no tenga "fundamento" o que no sea adaptativa. Más bien, simplemente sostengo que este comportamiento particular es, en este caso, irracional. Pocos, si es que alguno, de los participantes en este estudio elegirían conscientemente responder a la manipulación experimental (víctima determinada vs. víctima indeterminada) dando más a la víctima determinada. En otras palabras, este efecto experimental se habría visto reducido en una medida importante si no se hubiera eliminado por completo, es decir, si este experimento hubiera empleado un diseño centrado en el sujeto en lugar de un diseño centrado en la relación entre sujetos.

emoción. En un estudio de seguimiento que replica este efecto, los sujetos informaron sobre los niveles de simpatía y lástima que sentían por las víctimas determinadas/indeterminadas con las que estaban emparejados. Como era de esperar, sus niveles informados de simpatía y compasión estaban en concordancia con sus niveles de donación (comunicación personal).

Uno podría preguntarse si acaso este patrón se mantiene también afuera del laboratorio. Para averiguarlo, Small y Loewenstein llevaron a cabo un estudio posterior en el que las personas podían donar dinero a *Habitat for Humanity* con el fin de proporcionar un hogar a una familia necesitada, donde la familia podía estar determinada o por determinar. Tal como se predecía, por un lado, la donación media fue un veinticinco por ciento más alta en la condición familiar determinada y, por otro lado, la mediana en la condición familiar determinada fue el doble que la de la condición familiar indeterminada.

Y luego está la bebé Jessica. No podemos decir con certeza que los recursos se dirigieron a ella en lugar de a causas que podrían usar el dinero de manera más efectiva debido a las respuestas emocionales de las personas donantes (y no debido al razonamiento deontológico sobre derechos y deberes), pero la evidencia que existe sugiere que ese es el caso. Como podría haber dicho Josef Stalin, "la muerte de un individuo determinado es una tragedia; un millón de muertes indeterminadas es una estadística".

2.4 La ira y los enfoques deontológicos del castigo

Si bien los consecuencialistas y los deontólogos están de acuerdo en que el castigo por malas acciones es necesario e importante, discrepan profundamente sobre la justificación adecuada de dicho castigo. Consecuencialistas como Jeremy Bentham (1982) sostienen que el castigo se justifica únicamente por sus efectos beneficiosos futuros, principalmente a través de la disuasión y (en el caso del derecho penal) la contención de individuos peligrosos. Si bien pocos negarían que la prevención de daños futuros habilite una justificación legítima para el castigo, muchos creen que tales consideraciones pragmáticas no son las únicas razones legítimas para castigar, ni siquiera las principales. Deontólogos como Kant (2002), por ejemplo, argumentan que la justificación principal del castigo es la retribución, con el fin de darles a los malhechores lo que merecen en función de lo que han hecho, independientemente de si dicha retribución evitará futuras malas acciones.

Uno podría preguntarse, entonces, sobre la psicología del castigador típico. ¿La gente castiga, o respalda el castigo, por sus efectos beneficiosos o castiga porque está motivada a dar a la gente lo que merece en proporción a su "maldad interna" (para usar la frase de Kant)? (Carlsmith et al. 2002; Kant 2002). Varios estudios analizan esta cuestión y los resultados son consistentes. Las personas respaldan las justificaciones consecuencialistas y retributivistas del castigo en abstracto, pero en la práctica, o cuando se enfrentan a elecciones hipotéticas

más concretas, sus motivos parecen predominantemente retributivistas. A su vez, estas inclinaciones retributivistas parecen estar impulsadas emocionalmente: las personas castigan en proporción a la medida en que las transgresiones las enojan.

Primero, consideremos si los juicios punitivos son predominantemente consecuencialistas o deontológicos/retributivistas[8]. Jonathan Baron y sus colegas han realizado una serie de experimentos que demuestran que los juicios punitivos de las personas son, en su mayor parte, retributivistas más que consecuencialistas. En un estudio, Baron y Ritov (1993) presentaron a un grupo de personas casos hipotéticos de responsabilidad corporativa en los que se podía exigir a las corporaciones en cuestión que paguen multas. En uno de los casos, una corporación que fabrica vacunas está siendo demandada porque un niño murió como resultado de ser inoculado contra la influenza. Los sujetos recibieron múltiples versiones de este caso. En una versión, se estipuló que una multa tendría un efecto disuasorio positivo. Es decir, haría que la empresa produzca una vacuna más segura. En una versión diferente, se estipuló que una multa tendría un efecto "perverso". En lugar de hacer que la empresa busque generar una vacuna más segura, conllevaría que deje de fabricar este tipo de vacuna por completo. Se trata de un resultado indeseable, dado que la vacuna en cuestión hace más bien que daño y ninguna otra empresa sería capaz de hacerla. Los sujetos debían indicar si pensaban que una multa punitiva era apropiada para los dos casos o si la multa debería diferir entre ambos. La mayoría de los sujetos dijo que la multa no debería diferir en absoluto. Baron y Ritov lograron resultados similares utilizando una manipulación complementaria relativa a los efectos disuasorios sobre las decisiones de otras empresas. En un grupo experimental diferente, Baron y sus colegas encontraron una indiferencia similar hacia los factores consecuencialistas en respuesta a preguntas sobre el manejo de desechos peligrosos (Baron et al. 1993).

Los resultados de estos estudios son sorprendentes, sobre todo teniendo en cuenta que muchas personas consideran la disuasión de futuras decisiones perjudiciales como una razón principal, si no la razón principal, para imponer tales multas en el mundo real. También vale la pena destacar la fuerza de estos resultados. El hallazgo aquí no es simplemente que los juicios punitivos no concuerdan con el consecuencialismo, esto es, la opinión de que las consecuencias son, en última instancia, lo único que debería importar a los tomadores de decisiones. Mucho más que eso, parece que la mayoría de la gente no le da importancia alguna a factores que son de clara importancia consecuencialista, al menos en los contextos considerados.

[8] Estoy asumiendo que dentro del dominio del castigo, "deontológico" y "retribucionista" son efectivamente intercambiables, aunque sean conceptualmente distintos (por ejemplo, uno podría favorecer el castigo como un fin en sí mismo, pero de maneras impredecibles que desafían todas las reglas normativas). Hasta donde yo sé, todas las alternativas bien desarrolladas a las teorías consecuencialistas del castigo son, de una forma u otra, retributivas. Además, el retribucionismo está respaldado explícitamente por muchos deontólogos notables, incluido Kant (2002).

Si la gente no castiga por razones consecuencialistas, ¿qué los motiva entonces? En un estudio de Kahneman y colegas (1998), los sujetos respondieron a una serie de escenarios hipotéticos similares (por ejemplo, un caso de anemia debido a la exposición al benceno en el trabajo). Los sujetos valoraron cada escenario la medida en que la acción del acusado resultaba "indignante". También valoraron la medida en que el acusado en cada caso debía ser castigado. La correlación entre sus valoraciones promedio de indignación de estos escenarios y sus valoraciones promedio de castigo fue casi perfecta, con un coeficiente de correlación de Pearson (r) de .98. (un valor de 1 indica una correlación perfecta). A partir de los resultados, Kahneman y sus colegas concluyeron que la medida en que las personas desean que una empresa sea castigada por su comportamiento depende casi por completo de la medida en que las personas están emocionalmente indignadas por dicho comportamiento.

Carlsmith y sus colegas (2002) llevaron a cabo un conjunto similar de estudios destinados explícitamente a determinar si las personas castigan por 23 razones consecuencialistas o deontológicas. Aquí, como antes, a los sujetos se les presentó escenarios que involucraban un comportamiento moral y legalmente culpable, en este caso perpetrado por individuos en lugar de corporaciones. Al igual que en el caso anterior, se pidió a los sujetos que indicaran qué tan severo debería ser el castigo para cada persona, primero en términos abstractos ("nada severo" a "extremadamente severo") y luego en términos más concretos ("no culpable"/sin castigo de "cadena perpetua"). Los experimentadores diversificaron los escenarios de manera que garantizaran diferentes niveles de castigo, dependiendo de la justificación del castigo. Por ejemplo, una teoría consecuencialista del castigo considera que la tasa de detección asociada con un determinado tipo de delito y la publicidad asociada con un determinado tipo de condena son factores relevantes en la asignación de castigos. En otros términos, según los consecuencialistas, si un delito es difícil de detectar, entonces el castigo por ese delito debería ser más severo para así contrarrestar la tentación creada por el bajo riesgo de ser atrapado. Del mismo modo, si es probable que una condena obtenga mucha publicidad, un sistema de aplicación de la ley interesado en la disuasión debería aprovechar esta circunstancia "dando el ejemplo" con un castigo particularmente severo para el condenado en cuestión, obteniendo así el máximo de disuasión a través de un único castigado.

Los resultados fueron claros. Para el grupo experimental en su conjunto, cuando se manipularon las tasas de detección y los niveles de publicidad, no hubo cambios significativos en las recomendaciones de castigo. En otras palabras, la gente era generalmente indiferente a factores que, según los consecuencialistas, deberían importar, al menos hasta cierto punto. Esto sucedió a pesar del hecho de que Carlsmith et al., así como también otros (Weiner et al. 1997), habían encontrado que los sujetos expresaban fácilmente un tipo general de apoyo a los sistemas penales orientados a la disuasión y a las políticas corporativas.

En un estudio de seguimiento, se instruyó explícitamente a los sujetos para que adoptaran un enfoque consecuencialista, con la justificación consecuencialista expuesta expresamente y con controles de manipulación adicionales incluidos para garantizar que los sujetos pudiesen comprender los hechos relevantes. Aquí también los resultados fueron sorprendentes.

Cuando se les dijo que pensaran como consecuencialistas, los sujetos modificaron sus juicios, pero no de una manera genuinamente consecuencialista. En lugar de volverse selectivamente sensibles a los factores que aumentan los beneficios consecuencialistas del castigo, aumentaron indiscriminadamente el nivel de castigo en todos los casos, dando a los perpetradores el castigo que pensaban que merecían en función de las acciones que habían realizado, sumado a un poco más para favorecer el efecto disuatorio.

¿Qué motivó los juicios punitivos de estos sujetos? Aquí también, una parte importante de la respuesta parece ser «indignación». Los sujetos indicaron hasta qué punto estaban «moralmente indignados» por los delitos en cuestión y resultó que el grado de indignación moral en respuesta a un delito determinado era un buen indicador de la gravedad del castigo asignado al perpetrador (aunque el efecto aquí fue más débil que el observado en el estudio de Kahneman et al.) [9]. Además, un modelo de ecuación estructural de estos datos sugiere que los factores que tuvieron mayor efecto en los juicios de las personas sobre el castigo (gravedad del delito, presencia de circunstancias atenuantes) ejercieron su efecto a través de la "indignación moral".

Este estudio está en sintonía con la investigación de Small y Loewenstein sobre el "efecto víctima identificable" analizado en la sección anterior. Más recientemente, han documentado un efecto paralelo en el ámbito del castigo. En dicho experimento los sujetos realizaban un "juego de inversión" en el que los individuos reciben cierta cantidad de dinero que pueden elegir poner en un fondo de inversión colectiva. El juego permite a las personas elegir hasta qué punto jugarán de forma cooperativa, beneficiando al grupo a expensas de quien elige. Después del juego, los cooperadores tenían la oportunidad de castigar a los jugadores egoístas haciéndoles perder dinero, pero los cooperadores que castigaban tenían que pagar por ello. Como antes, la manipulación crucial fue entre individuos determinados e indeterminados, en este caso los jugadores egoístas. A algunos sujetos se les preguntó "¿Qué tanto preferirías castigar al sujeto n°4 no cooperador?" y a otros sujetos "¿Qué tanto preferirías castigar al sujeto no cooperador cuya carta te tocará?". De acuerdo con los resultados anteriores, el castigo

[9] Al interpretar los resultados de estos dos estudios de "indignación" y castigo, surgen algunas complicaciones. No está claro si la escala de "indignación" utilizada por Kahneman et al. obtiene un informe subjetivo del estado emocional del sujeto o un juicio normativo sobre el comportamiento del acusado. Un escéptico podría decir que la llamada escala de "indignación" es en realidad solo una escala de la gravedad moral general del delito, lo que, como era de esperar, se correlaciona con la medida en que las personas piensan que merece un castigo. El estudio de Carlsmith et al. aborda esta preocupación (aunque no intencionalmente) y sugiere que puede tener alguna validez. La medida de indignación utilizada en dicho estudio pide explícitamente un informe subjetivo: "¿Qué tan indignado a nivel moral estaba usted por esta ofensa?". Y, tal vez como resultado de este cambio de táctica, la conexión entre "indignación" y juicio punitivo se debilita de 'casi perfecto' a 'bastante fuerte'. Tenga en cuenta también que al elegir una palabra fuerte como "indignación" en un estudio de delitos hipotéticos bastante leves, los experimentadores pueden haber puesto la vara demasiado alta para los informes subjetivos, debilitando así sus resultados.

promedio fue casi el doble para el grupo determinado y, una vez más, los informes de los sujetos sobre sus respuestas emocionales (en este caso una medida compuesta de ira y culpa) resultaron compatibles con su comportamiento (Small y Loewenstein, en prensa).

Estudios recientes de neuroimagen también sugieren que el deseo de castigar está impulsado emocionalmente. Alan Sanfey, Jim Rilling y sus colegas (2003) realizaron un estudio de imágenes cerebrales del juego del ultimátum, con el fin de estudiar las bases neuronales del sentido de justicia. El juego del ultimátum funciona de la siguiente manera. Hay una suma de dinero, digamos $10, y el primer jugador (el proponente) hace una propuesta sobre cómo dividirlo entre el otro jugador y él. El segundo jugador, el respondedor, puede aceptar la oferta, en cuyo caso el dinero se divide según lo propuesto, o rechazar la oferta, en cuyo caso nadie recibe nada. Los proponentes suelen hacer ofertas que son justas (es decir, divididas al cincuenta por ciento) o cercanas a lo justo, y los que responden tienden a rechazar ofertas que son más que un poco injustas. En otras palabras, los respondedores generalmente pagarán por el privilegio de castigar a los proponentes injustos, incluso cuando se trata de un juego de una sola vez. ¿Por qué las personas hacen esto?

La respuesta, una vez más, involucra emoción. Los experimentadores encontraron que las ofertas injustas, en comparación con las ofertas justas, producían una mayor actividad en la ínsula anterior, una región del cerebro asociada con la ira, el disgusto y la excitación autónoma. A su vez, los niveles promedio de actividad de la ínsula de los individuos se correlacionaron positivamente con el porcentaje de ofertas que rechazaban y fueron más débiles para los ensayos en los que el sujeto creía que la oferta injusta había sido hecha por un programa de computadora en lugar de por una persona real. Por supuesto, es concebible que las personas estuvieran castigando a los proponentes injustos en un intento de disuadirlos de ser injustos con los demás en el futuro, pero dada la constancia con la que las personas son insensibles a las manipulaciones que modulan los efectos disuasorios del castigo parece poco probable. En cambio, resulta mucho más probable que la gente haya infligido un castigo buscando su beneficio propio. Una vez más, parece ser que esta tendencia retributivista está impulsada emocionalmente. Un estudio de neuroimagen más reciente sobre el castigo en respuesta a violaciones de la confianza arroja una conclusión similar (de Quervain et al. 2004). En dicho estudio, la extensión del castigo se correlacionó con el nivel de actividad en el núcleo caudado, una región del cerebro asociada con la emoción y relacionada más específicamente con la motivación y la recompensa.

Cuando a las personas se les pregunta de manera general y abstracta por qué tiene sentido castigar, los argumentos consecuencialistas son prominentes (Carlsmith et al. 2002; Weiner et al. 1997). Ahora bien, cuando se les presentan casos más concretos que involucran a individuos específicos que cometen delitos específicos, sus juicios resultan en gran medida, y en muchos casos completamente, insensibles a los factores que afectan las consecuencias del castigo. Esto es así incluso cuando se enfatiza la lógica consecuencialista al considerar estos factores y cuando se instruye explícitamente a las personas a pensar como consecuencialistas. Parece, entonces, que el pensamiento consecuencialista juega un papel insignificante en el jui-

cio punitivo de sentido común y que el juicio punitivo de sentido común es casi enteramente retributivista/deontológico, siempre y cuando el asunto sea lo suficientemente concreto. La evidencia disponible, además, tanto de evaluaciones como de datos de neuroimágenes, sugiere que los juicios punitivos deontológicos/retributivistas son predominantemente emocionales, impulsados por sentimientos de ira o "indignación".

2.5 La emoción y la condena moral de las acciones inofensivas

Según los consecuencialistas, las acciones están mal debido a sus consecuencias dañinas. En contraste, los deontólogos, junto con muchos moralistas de sentido común, podrían condenar acciones que no causan daño en ningún sentido ordinario. Por ejemplo, un deontólogo probablemente diría que no está bien romper promesas, independientemente de si hacerlo pudiese tener consecuencias perjudiciales. Jonathan Haidt (Haidt et al. 1993) ha realizado una serie de estudios sobre la formulación de juicios morales en respuesta a acciones inofensivas. De este trabajo surgen dos temas relevantes para la presente discusión. En primer lugar, la condena moral de una acción inofensiva parece estar impulsada por la emoción. En segundo lugar, la experiencia que fomenta un enfoque más "cognitivo" de la toma de decisiones morales tiende a hacer que las personas estén menos dispuestas a condenar acciones inofensivas.

Haidt y dos colegas brasileños realizaron un estudio transcultural del juicio moral utilizando un gran conjunto de sujetos que variaban en estatus socioeconómico (SES), nacionalidad (brasileña y estadounidense) y edad (niños y adultos). A los sujetos se les presentaron una serie de escenarios que involucraban acciones inofensivas y moralmente cuestionables:

1. Un hijo le promete a su madre moribunda que visitará su tumba todas las semanas después de su muerte, pero luego no lo hace porque está ocupado.
2. Una mujer usa una vieja bandera estadounidense/brasileña para limpiar el baño.
3. Una familia se come a su perro después de haber sido asesinado accidentalmente por un automóvil.
4. Un hermano y una hermana se besan en los labios.
5. Un hombre se masturba usando un pollo muerto, antes de cocinarlo y comérselo.

Luego de presentarles estos casos, los sujetos respondieron las siguientes preguntas: ¿Esta acción es incorrecta? Si es así, ¿por qué? ¿Esta acción daña a alguien? Si vieras a alguien hacer esto, ¿te molestaría? ¿Alguien que hace esto debe ser detenido o castigado? Si hacer esto es una costumbre en algún país extranjero, ¿es incorrecta dicha costumbre?

Cuando las personas dicen que ciertas acciones están mal, ¿por qué lo dicen? Una hipótesis es que estas acciones se perciben como dañinas, lo sean o no (Turiel et al. 1987): besar en la boca a un hermano podría causar un daño psicológico; masturbarse con una gallina podría propagar enfermedades, etc. Si esta hipótesis es correcta, entonces podríamos esperar que las

respuestas por parte de la gente a la pregunta "¿Esta acción daña a alguien?" se correlacionen con el grado de condena moral, como lo indican las respuestas afirmativas a las preguntas: "¿Esto está mal?" "¿Esta persona debe ser detenida o castigada?" "¿Está mal si es la costumbre local?" Alternativamente, si en este tipo de casos son las emociones las que impulsan la condena moral, entonces podríamos esperar que la respuesta a la pregunta "Si vieras esto, ¿te molestaría?" permitiría predecir mejor las respuestas a las preguntas morales anteriores. Como se esperaba, Haidt y sus colegas encontraron que una respuesta afirmativa a la pregunta "¿Acaso le molestaría esto?" era un mejor predictor de la condena moral que una respuesta afirmativa a la pregunta sobre el daño[10].

Igualmente interesantes fueron las diferencias entre los diferentes grupos. Primero, los sujetos de alto estatus socioeconómico y educacional (ESE) en Filadelfia y Brasil fueron mucho menos condenatorios que sus contrapartes de bajo ESE, tanto incluso que los grupos de alto ESE en Filadelfia y Brasil se parecían más entre sí de lo que se parecían a sus vecinos de bajo ESE. En segundo lugar, la gente de las ciudades menos "occidentalizadas" tenía una tendencia a condenar más[11]. En tercer lugar, los niños de ambos lugares tendían a ser más condenatorios que los adultos. En otras palabras, la educación (ESE), la occidentalización y el nivel de crecimiento, en respuesta a los escenarios aquí utilizados, se asociaron con juicios más consecuencialistas. Estos tres hallazgos tienen sentido a la luz del modelo de juicio moral que hemos estado desarrollando, según el cual mientras que las respuestas emocionales intuitivas impulsan intuiciones morales prevalentes, los procesos de control "cognitivos" solo a veces logran dominarlas. La educación es, en gran medida, el desarrollo de las propias capacidades "cognitivas", es decir, aprender a pensar de formas abstractas, esforzadas y, a menudo, no intuitivas o contraintuitivas. El factor de occidentalización está estrechamente relacionado con esto. Si bien los occidentales pueden no estar más "cognitivamente" desarrollados que los miembros de otras culturas, la tradición occidental adopta lo que es, desde una perspectiva antropológica, un enfoque peculiarmente "cognitivo" de la moral. En comparación con miembros de otras culturas, los occidentales son más propensos a defender y justificar sus creencias y valores morales en términos abstractos (Rozin, comunicación personal). Además, la cultura occidental tiende a ser más pluralista que otras culturas, valorando explícitamente múltiples perspectivas e, incluso, una conciencia intelectual de que existen perspectivas alternativas. Finalmente, la capacidad de "control cognitivo" continúa desarrollándose durante

[10] Este resultado, sin embargo, solo se mantuvo para los subgrupos que sostuvieron la mayoría de las condenas. Los sujetos que se mostraron más reacios a condenar violaciones inofensivas (principalmente occidentales educados y de nivel socioeconómico alto), encontraron daño donde otros no lo hicieron y lo citaron como motivo de condena, un efecto que Haidt ha documentado en otro lugar y que ha denominado "desconcierto moral" (Haidt, Bjorklund y Murphy 2000).

[11] "Occidentalización" se refiere al "grado en que cada una de las tres ciudades [Filadelfia y dos ciudades brasileñas, Porto Alegre y Recife] tiene una vida cultural y simbólica basada en las tradiciones europeas, incluida una estructura política democrática y una economía industrializada" (Haidt, Koller y Dias 1993, 615). Filadelfia está más occidentalizada que Porto Alegre, última la cual, a su vez, está más occidentalizada que Recife.

la adolescencia (Anderson et al. 2001; Paus et al. 1999). Si bien los niños, al igual que los adultos, son muy buenos para sentir emociones como la ira, la simpatía, el disgusto, etc., no son muy buenos para controlar su comportamiento cuando experimentan tales sentimientos (Steinberg y Scott, 2003). Por tanto, al igual que lo que ocurría con lo desarrollado más arriba, parece haber un vínculo entre "cognición" y juicio consecuencialista.

En este estudio, la conexión entre la renuencia a condenar y el consecuencialismo es bastante sencilla: los consecuencialistas no condenan acciones inofensivas[12]. La conexión entre la tendencia a condenar las acciones inofensivas y la deontología es, sin embargo, menos directa y más cuestionable. No es obvio, por ejemplo, que los deontólogos sean más propensos que los consecuencialistas a condenar la profanación de banderas o el comerse el perro de la familia. Dudas similares se aplican al caso de los hermanos que se besan y al hombre que se masturba con una gallina muerta, aunque vale la pena señalar que Kant argumentó que el incesto, la masturbación, la bestialidad y casi todas las demás formas de experimentación sexual van en contra de la ley moral (Kant 1930; Kant 1994). El caso de la promesa rota, sin embargo, es "deontología de núcleo". Por supuesto, no todos los deontólogos condenarían el incumplimiento inofensivo de una promesa a la madre fallecida, pero cualquiera que condene tal comportamiento (sin apelar de alguna manera a las consecuencias) está exhibiendo un comportamiento característicamente deontológico[13]. A la luz de este punto, vale la pena examinar el caso en cuestión un poco más de cerca, dado que, de hecho, resulta que encaja bastante bien en el patrón de las diferencias entre grupos. Entre los adultos con ESE alto, mientras que el porcentaje de sujetos de cada ciudad que sostuvieron que esta acción debería ser detenida o castigada osciló entre el 3% y el 7%, el porcentaje de adultos con ESE bajo que dijeron lo mismo osciló entre el 20% (Filadelfia) y el 57%. % (Recife, Brasil). Del mismo modo, mientras que entre los adultos con un ESE alto el porcentaje que afirmó que se trataba de un comportamiento incorrecto incluso si fuese la costumbre local osciló entre el 20% y el 28%, los porcentajes correspondientes para los sujetos con ESE bajo osciló entre el 40% y el 87%. La tendencia a condenar este comportamiento también disminuyó con la occidentalización y, dentro de cada grupo, los niños estaban más dispuestos a condenarlo que los adultos (si quieres que alguien visite tu tumba cuando estés muerto, nadie le gana a los niños pobres de Recife, Brasil; el 97% apoya castigar/detener a las personas que no cumplen las promesas de visitar la tumba y el 100% condena las culturas en las que se acostumbra a ello). Por lo tanto, el argumento anterior que conecta la "cognición" y el consecuencialismo se aplica específicamente al caso en el que la condena moral es más característicamente deontológica. Haidt y colegas no proporcionaron datos sobre la pregunta "¿Acaso le molestaría esto?" para este

[12] Un consecuencialista podría estar a favor de la prohibición de una clase de acciones, algunas de las cuales podrían no ser dañinas, si la prohibición produce las mejores consecuencias disponibles. Del mismo modo, un consecuencialista podría pretender condenar una acción (o condenar públicamente mientras que en privado se abstiene de condenar), si esta condena pública se considerara beneficiosa.

[13] Se pidió a los sujetos que justificaran sus respuestas. Las justificaciones típicas para condenar esta acción no apelaron a las consecuencias, sino que simplemente declararon que estaba mal romper una promesa.

caso específicamente, pero el hecho de que no sea una excepción al patrón "cognitivo" general (menos condena en presencia de factores que estimulan la "cognición") sugiere que es poco probable que sea una excepción al patrón general relacionado con las emociones (condena correlacionada con emociones negativas).

Una evidencia más poderosa y directa del papel de la emoción en la condena de las violaciones morales inofensivas proviene de dos estudios más recientes. En el primero de ellos, Thalia Wheatley y Jonathan Haidt (Wheatley y Haidt 2005) generaron en un grupo de individuos hipnotizables la sugestión post-hipnótica de sentir una punzada de disgusto al leer la palabra "a menudo" (y luego el olvido de que recibieron esta sugestión). Otro grupo de sujetos (también individuos hipnotizables) recibió el mismo tratamiento, excepto que en este caso fue sensibilizado con la palabra "realizar". A continuación, a los sujetos se les presentó distintos escenarios, algunos de los cuales no implicaban ningún daño. En un escenario, por ejemplo, primos segundos tienen una relación sexual en la que "realizan/a menudo hacen viajes de fin de semana a hoteles románticos en las montañas". Los sujetos que recibieron la sugestión post-hipnótica correspondiente (es decir, leyeron la palabra a la que estaban hipnóticamente sensibilizados), a diferencia de los individuos del otro grupo, juzgaron que las acciones de esta pareja eran más moralmente incorrectas.

En un segundo experimento, Wheatley y Haidt utilizaron un escenario en el que la persona descrita no hacía nada malo en absoluto. Se trataba del caso de un representante de un consejo estudiantil que "a menudo elige" (o "trata de realizar") temas de amplio interés para discutirlos. Muchos sujetos que recibieron una sugestión post-hipnótica coincidente indicaron que su comportamiento era un poco incorrecto y dos, en particular, le dieron una calificación de incorrección alta. Los sujetos decían cosas como "Parece que está tramando algo", "Parece tan extraño y repugnante" y "No sé [por qué está mal], simplemente es así". Nuevamente, vemos que las emociones hacen que las personas tiendan a conclusiones no consecuencialistas.

En un estudio más reciente (no publicado), Simone Schnall, Jonathan Haidt y Gerald Clore (Schnall et al. 2004) manipularon los sentimientos de disgusto de un grupo de sujetos, en este caso no con hipnosis sino sentándolos, mientras llenaban sus cuestionarios, en un escritorio repugnante (estaba manchado y pegajoso, ubicado cerca de un bote de basura desbordado que contenía cajas de pizza usadas y pañuelos de papel de aspecto sucio, etc.). Los sujetos respondieron a una serie de escenarios que evaluaban el juicio moral, incluyendo variaciones en los casos de la masturbación y de comerse al perro anteriormente mencionados. Aquí, como antes, la manipulación del disgusto hizo que las personas fueran más propensas a condenar estas acciones, aunque solo en el caso de sujetos que fueron calificados como altamente sensibles a sus propios estados corporales.

2.6 Dos patrones de juicio moral

Los experimentos llevados a cabo por Greene et al., Small y Loewenstein, Baron et al., Kahneman et al., Carlsmith et al., Sanfey et al., De Quervain et al. y Haidt et al., proporcionan de manera conjunta múltiples evidencias independientes de que mientras que los patrones deontológicos del juicio moral son impulsados por respuestas emocionales, los juicios consecuencialistas son impulsados por procesos "cognitivos". Cualquiera de los resultados e interpretaciones descritos anteriormente puede ser cuestionado, pero la evidencia convergente reunida aquí constituye un caso decente para la asociación entre deontología y emoción, especialmente porque, al menos hasta donde conozco, no hay evidencia empírica de lo contrario. Por supuesto, los deontólogos pueden considerarse a sí mismos y a sus mentes como excepciones a los patrones psicológicos convergentes y estadísticamente significativos identificados en estos estudios, sin embargo, en mi opinión, la carga de la prueba recae sobre ellos a la hora de demostrar que son psicológicamente excepcionales de una manera que preserve sus autoconcepciones.

¿Por qué van juntas la deontología y la emoción? Creo que la respuesta tiene dos partes. En primer lugar, la emoción moral proporciona una solución natural a ciertos problemas generados por la vida social. En segundo lugar, la filosofía deontológica proporciona una interpretación «cognitiva» natural de la emoción moral. Consideremos cada una de estas afirmaciones.

En primer lugar, ¿por qué emociones morales? En las últimas décadas se han presentado muchas explicaciones plausibles y complementarias, a partir de las cuales parece estar surgiendo un consenso general: las emociones más relevantes para la moral existen porque motivan comportamientos que ayudan a los individuos a difundir sus genes dentro de un contexto social. La teoría de la selección de parentesco explica por qué los individuos tienden a preocuparse por el bienestar de aquellos con quienes están estrechamente relacionados (Hamilton 1964): un individuo puede propagar sus propios genes ayudando a familiares cercanos a difundir los suyos, debido a que, justamente, los parientes cercanos comparten una alta proporción de sus genes. En relación con lo anterior, la teoría del altruismo recíproco explica la existencia de una forma más amplia de altruismo: los individuos sin parentesco genético pueden beneficiarse de ser amables entre sí, siempre y cuando sean capaces de identificar quién estaría dispuesto a devolver dicha amabilidad (Trivers 1971). Las teorías evolutivas más recientes del altruismo intentan explicar la evolución de la "reciprocidad fuerte", una tendencia más amplia a recompensar el comportamiento cooperativo y a castigar el comportamiento no cooperativo, incluso en contextos en los que las condiciones necesarias para la selección de parentesco (relaciones genéticas detectables) y el altruismo recíproco (disposiciones cooperativas detectables) no se cumplen (Bowles y Gintis 2004; Fehr y Rockenbach 2004; Gintis 2000). Estas teorías buscan explicar la tendencia humana generalizada a participar en comportamientos cooperativos (p. ej., ayudar a los demás y hablar con honestidad) y evitar comportamientos no cooperativos (p. ej., herir a otros y mentir), incluso cuando los familiares y colaboradores cercanos no están involucrados. Además, dichas teorías pretenden darle explicación al "castigo altruista", la disposición de las personas a castigar el comportamiento antisocial incluso

aun cuando no podrían esperar beneficiarse de ello (Boyd et al. 2003; Fehr y Gachter 2002; Fehr y Rockenbach 2004). Otras teorías evolutivas intentan dar sentido a otros aspectos de la moral. Por ejemplo, el tabú del incesto puede explicarse como un mecanismo para evitar defectos de nacimiento, los cuales resultan con mayor probabilidad de apareamientos entre parientes cercanos (Lieberman et al. 2003). Finalmente, el campo emergente de la evolución cultural promete explicar cómo las normas morales (y las prácticas culturales de una manera más amplia) se desarrollan y difunden (Richerson y Boyd 2005).

Estas explicaciones evolutivas de los fenómenos morales han recibido mucha atención en los últimos años (Pinker 2002; Sober y Wilson 1998; Wright 1994) y, por lo tanto, no los detallaré aquí. Simplemente asumiré que la idea general de estas teorías es correcta, a saber, que nuestras disposiciones morales más básicas son adaptaciones evolutivas que surgieron en respuesta a las demandas y oportunidades creadas por la vida social. La pregunta pertinente aquí es la relacionada con la implementación psicológica de estas disposiciones. ¿Por qué nuestro comportamiento moral adaptativo debería estar impulsado por emociones morales en lugar de por otra cosa, como el razonamiento moral? Creo que la respuesta es que mientras que las emociones son respuestas muy confiables, rápidas y eficientes a situaciones recurrentes, el razonamiento es poco confiable, lento e ineficiente en tales contextos (véase Sober y Wilson [1998, cap. 10] sobre emociones altruistas frente al razonamiento hedonista).

La naturaleza no le deja la tarea de descubrir que la ingestión de grasas y proteínas es propicia para nuestra supervivencia a nuestro poder de razonamiento. Más bien, nos da hambre y una sensación intuitiva de que cosas como la carne y la fruta lo satisfarán. La naturaleza no nos deja a nosotros la tarea de descubrir que los seres humanos son compañeros más adecuados que los babuinos. En cambio, nos dota de una psicología que hace que ciertos humanos nos parezcan parejas sexuales muy atractivas y que los babuinos nos parezcan terriblemente desagradables en este sentido. Por último, la naturaleza no nos deja a nosotros la tarea de darnos cuenta de que salvar a un niño que se está ahogando es algo bueno. En cambio, nos dota de un poderoso "sentido moral" que nos obliga a participar en este tipo de comportamiento (en las circunstancias adecuadas). En resumen, cuando la naturaleza necesita que un trabajo comportamental sea concretado, lo hace, siempre que puede, con la intuición y la emoción.

Por lo tanto, desde un punto de vista evolutivo, no es de extrañar que las disposiciones morales hayan evolucionado ni tampoco que estas se implementen emocionalmente. Ahora bien, yendo a la segunda parte de la explicación. ¿Por qué la existencia de emociones morales favorecería la existencia de una filosofía deontológica? Para responder a esta pregunta debemos apelar al hecho bien documentado de que los humanos son, en general, explicadores y justificadores irreprimibles de su propio comportamiento. Repetidamente, los psicólogos han descubierto que cuando las personas no saben por qué están haciendo lo que están haciendo, simplemente inventan una historia que suena plausible (Haidt 2001; Wilson 2002).

Recordemos, por ejemplo, el experimento de las medias mencionado al principio. Los sujetos no sabían que se sentían atraídos por los elementos del lado derecho de la pantalla,

pero cuando se les pedía que explicaran sus preferencias inventaban explicaciones alternativas perfectamente racionales (Nisbett y Wilson 1977). En un experimento similar, Nisbett y Wilson (1977) indujeron a un grupo de sujetos a preferir el detergente para ropa *Marea*, condicionándolos con pares de palabras como "océano-luna" en una prueba de memoria anterior. Cuando los sujetos explicaron sus preferencias, dijeron cosas como "*Marea* es el detergente más conocido", o "Mi madre usa *Marea*" o "Me gusta la caja de *Marea*". En un experimento inicial de Maier (Maier 1931; Nisbett y Wilson 1977), los sujetos tenían que encontrar una manera de atar dos cables que colgaban del techo. Se trataba de una tarea desafiante, ya que los cables estaban demasiado separados como para poder alcanzarlos simultáneamente. La solución era atar un objeto pesado a una de las cuerdas para que pudiera balancearse como un péndulo. El sujeto podría agarrarse a un cable mientras esperaba que el otro se balancee hacia su alcance. Maier ayudaba a sus sujetos a resolver este problema dándoles una pista sutil. Mientras caminaba por la habitación, ponía uno de los cables en movimiento de manera casual. Los sujetos que fueron ayudados por esta pista, no obstante, desconocían su influencia. En cambio, atribuían su intuición a una señal diferente y más llamativa (Maier girando una pesa en una cuerda), a pesar de que se demostró que esta pista era inútil en otras versiones del experimento. En un experimento similar, Dutton y Aron (Dutton y Aron 1974; Wilson 2002) hicieron que un grupo de sujetos masculinos cruzara un aterrador puente que se extendía a lo largo de un profundo barranco, después de lo cual se encontraba con una atractiva experimentadora. El grupo control descansaba en un banco antes de encontrarse con la experimentadora. Hallaron que aquellos que acababan de cruzar el puente aterrador, con las palmas sudorosas y el corazón latiendo con fuerza, tenían más del doble de probabilidades que los sujetos del grupo control de llamar más tarde a la experimentadora para pedirle una cita. Estos individuos (muchos de ellos, al menos) interpretaron su mayor excitación fisiológica como una mayor atracción por la mujer que habían conocido.

La tendencia hacia la racionalización post-hoc se revela a menudo en estudios de personas con condiciones mentales inusuales. Los pacientes con amnesia de Korsakoff y trastornos de la memoria relacionados son propensos a la "confabulación". Es decir, intentan disimular sus déficits de memoria construyendo elaboradas narraciones sobre sus historias personales, generalmente expresadas con gran confianza y sin conciencia aparente de que están inventando. Por ejemplo, a un paciente confabulador sentado cerca de un aire acondicionado se le preguntó si sabía dónde estaba. Él respondió que estaba en una planta de aire acondicionado. Cuando se le señaló que vestía pijama, dijo: "Los guardo en mi auto y pronto me cambiaré a mi ropa de trabajo" (Stuss et al. 1978). Del mismo modo, los individuos que actúan bajo sugestión poshipnótica a veces explican sus comportamientos en términos elaboradamente racionales. En un caso, a un sujeto hipnotizado se le indicó que, al percibir una señal arbitraria, colocara una pantalla de lámpara en la cabeza de otra persona. Hizo lo que le indicaron, pero cuando se le pidió que explicara por qué hizo lo que hizo, no se refirió a la sugerencia post-hipnótica o a la señal: "Bueno, te lo diré. Suena raro, pero es solo un pequeño experimento psicológico. Estuve leyendo sobre psicología del humor y pensé que me gustaría ver cómo reaccionaban

ante una broma de muy mal gusto" (Estabrooks 1943; Wilson 2002). Quizás el ejemplo más sorprendente de este tipo de racionalización post-hoc provenga de estudios de pacientes con el cerebro dividido, a saber, personas en las que no existe una comunicación neuronal directa entre los hemisferios cerebrales. En un estudio se le mostró al hemisferio derecho de un paciente una escena donde se veía nieve y se le indicó que seleccionara una imagen coincidente. Usando su mano izquierda, la mano controlada por el hemisferio derecho, seleccionó la imagen de una pala. Al mismo tiempo, a su hemisferio izquierdo, el hemisferio dominante para el lenguaje, se le mostró una imagen de una pata de pollo. Al paciente se le preguntó verbalmente por qué eligió la pala con la mano izquierda. Él respondió: "Vi una pata y elegí un pollo, y para limpiar el gallinero hace falta una pala" (Gazzaniga y Le Doux 1978; Wilson 2002). Gazzaniga y LeDoux argumentan que este tipo de confabulaciones no son exclusivas de los pacientes con cerebro dividido, es decir, que esta tendencia no se creó cuando se cortaron las líneas de comunicación en su cerebro. Más bien, argumentan, todos somos una suerte de confabuladores. Respondemos a los efectos conscientes de nuestros procesos perceptivos, mnemónicos y emocionales inconscientes dándoles forma a través de una narrativa racionalmente sensible, pero sin darnos cuenta de que lo estamos haciendo. Esta tendencia generalizada a la racionalización solo se revela en experimentos cuidadosamente controlados en los que los *inputs* psicológicos y los *outputs* conductuales pueden ser monitoreados cuidadosamente, o en estudios de individuos anormales que se ven obligados a construir una narrativa plausible a partir de escasa materia prima.

Ahora ya estamos listos para ir atando cabos. ¿Qué debemos esperar de las criaturas que (1) exhiben un comportamiento social/moral impulsado en gran parte por respuestas emocionales intuitivas y (2) que son propensas a la racionalización de sus comportamientos? Creo que la respuesta es la filosofía moral deontológica. ¿Qué sucede cuando nos imaginamos empujando al hombre grande por el puente peatonal? Si estoy en lo cierto, nos surge una respuesta emocional intuitiva que dice "¡No!". Esta voz negativa puede ser anulada, por supuesto, pero en lo que respecta a la voz en sí misma, no hay posibilidad de negociación. Ya sea que en última instancia podamos o no justificar el empujar al hombre del puente, siempre nos sentiremos mal. Y qué mejor manera de expresar ese sentimiento de absoluta injusticia no negociable que a través del concepto deontológico más central, el concepto de un derecho: no se le puede empujar a la muerte porque eso sería una violación de sus derechos. Del mismo modo, no puedes dejar que el bebé se ahogue porque tienes el deber de salvarlo.

La deontología, entonces, es un tipo de confabulación moral. Tenemos sentimientos fuertes que nos dicen en términos claros e inciertos que algunas cosas simplemente no se pueden hacer y que otras simplemente deben hacerse. Pero no es obvio cómo darle sentido a estos sentimientos, por lo que, con la ayuda de algunos filósofos especialmente creativos, inventamos una narración racionalmente atractiva: existen estas cosas llamadas "derechos" que la gente tiene y cuando alguien tiene un derecho no puedes hacer nada para quitárselo. No importa si el tipo del puente está llegando al final de su vida natural o si hay siete personas en las vías en lugar de cinco. Si el hombre tiene un derecho, entonces el hombre tiene un

derecho. Como dijo John Rawls (Rawls, 1971, 3-4), "Cada persona posee una inviolabilidad fundada en la justicia que ni siquiera el bienestar de la sociedad en conjunto puede atropellar" y " En una sociedad justa, las libertades de la igualdad de ciudadanía se dan por establecidas definitivamente; los derechos asegurados por la justicia no están sujetos a regateos políticos ni al cálculo de intereses sociales"[14]. Estas son líneas que nos hacen aplaudir porque tienen sentido a nivel emocional. La deontología, creo, es una expresión "cognitiva" natural de nuestras emociones morales más profundas.

Esta hipótesis plantea una pregunta adicional. ¿Por qué solo deontología? ¿Por qué no suponer que toda la filosofía moral, incluso todo el razonamiento moral, es una racionalización de las emociones morales? (esta es la forma fuerte del punto de vista defendido por Jonathan Haidt (2001), cuyo argumento es la referencia del argumento aquí presentado[15]). La respuesta, creo, es que el juicio moral consecuencialista no está impulsado por la emoción, o al menos no por el tipo de emoción de tipo 'alarma' que sí impulsa al juicio deontológico. La evidencia presentada anteriormente apoya esta hipótesis: lo que sugiere es que el juicio consecuencialista es menos emocional y más "cognitivo", pero no explica por qué debería ser así. Argumenté anteriormente que existe un mapeo natural entre el contenido de la filosofía deontológica y las propiedades funcionales de las emociones de tipo alarma. Asimismo, creo que existe un mapeo natural entre el contenido de la filosofía consecuencialista y las propiedades funcionales de los procesos "cognitivos". De hecho, considero que el consecuencialismo es inherentemente "cognitivo" y que no podría implementarse de otra manera.

El consecuencialismo es, por su propia naturaleza, sistemático y agregativo. Su objetivo es tener en cuenta casi todo y garantiza que casi todo es negociable. Toda toma de decisiones consecuencialista se caracteriza por equilibrar preocupaciones en competencia, teniendo en cuenta tanta información como sea prácticamente posible. Solo en ejemplos hipotéticos en los que " todo el resto se mantiene constante", el consecuencialismo ofrece respuestas claras. Para el consecuencialismo de la vida real, todo es un complejo juego de adivinanzas y todos los juicios pueden ser revisados a la luz de detalles adicionales. No hay claridad moral en el pensamiento moral consecuencialista, con sus aproximaciones y suposiciones simplificadoras. Es fundamentalmente actuarial.

Recordemos la definición de "cognitivo" propuesta anteriormente: las representaciones "cognitivas" son representaciones inherentemente neutrales que, a diferencia de las representaciones emocionales, no desencadenan automáticamente respuestas o disposiciones conductuales particulares. Una vez más, la ventaja de tener tales representaciones neutrales es que pueden mezclarse y combinarse según la especificidad de la situación, sin tironear al agente en múltiples direcciones de comportamiento al mismo tiempo, lo cual habilita un compor-

[14] N. del T.: Para la traducción de estas líneas se utilizó la de Rawls, J. (2006). *Teoría de Justicia* (pp. 17). México: Fondo de Cultura Económica.

[15] Haidt (2001), sin embargo, considera que los filósofos pueden ser excepcionales en el sentido de que realmente razonan para llegar a conclusiones morales (Kuhn 1991).

tamiento altamente flexible. Son precisamente estos tipos de representaciones los que precisa un consecuencialista para emitir un juicio basado en el acopio de información, es decir, uno que tenga en cuenta todos los factores relevantes: "¿Está bien empujar al tipo por el puente si está a punto de curar el cáncer?", "¿Está bien salir a comer sushi cuando el dinero extra podría usarse para promover la educación sanitaria en África?" Y así. Los deontólogos podrían descartar este tipo de preguntas complicadas y específicas de ciertas situaciones, pero los consecuencialistas no pueden, razón por la cual sostengo que el consecuencialismo es ineludiblemente "cognitivo".

Algunas aclaraciones. En primer lugar, no estoy afirmando que el juicio consecuencialista se encuentra exento de la influencia de las emociones. Por el contrario, me inclino a estar de acuerdo con Hume (Hume 1978) respecto de que todo juicio moral tiene algún componente afectivo. De hecho, sospecho que la ponderación consecuencialista de daños y beneficios es un proceso emocional. Ahora bien, si estoy en lo cierto, dos cosas distinguen este tipo de proceso de los asociados con la deontología. En primer lugar, se trata, como dije anteriormente, de un proceso en el que se sopesa información disponible y no uno de tipo "alarma". Los tipos de emociones hipotéticas que están involucradas dicen: "Esto y aquello importa este tanto. Téngalo en cuenta". Por su parte, las emociones hipotéticas para impulsar el juicio deontológico son mucho menos sutiles. Son, como afirmé, señales de alarma que emiten comandos simples: "¡No lo haga!" o "¡Debe hacerlo!". Si bien estos comandos se pueden anular, están diseñados para dominar la decisión en lugar de simplemente influir en ella.

En segundo lugar, no estoy afirmando que el juicio deontológico no pueda ser "cognitivo". De hecho, creo que a veces lo es (véase más abajo). Más bien, mi hipótesis reside en que el juicio deontológico es afectivo en su esencia, mientras que el juicio consecuencialista es ineludiblemente "cognitivo". Se podría, en principio, hacer un juicio característicamente deontológico pensando explícitamente en el imperativo categórico y en si la acción en cuestión se basa en una máxima que podría servir como ley universal. Si uno hiciera eso, entonces el proceso psicológico sería "cognitivo". Lo que propongo, sin embargo, es que no es así como se tiende a llegar a conclusiones característicamente deontológicas, sino que, en cambio, tienden a alcanzarse sobre la base de respuestas emocionales. Esto contrasta con los juicios consecuencialistas que, según mi hipótesis, simplemente no pueden implementarse de manera intuitiva y emocional. La única forma de llegar a un juicio distintivamente consecuencialista (es decir, uno que no coincide con un juicio deontológico) es a través del razonamiento consecuencialista de costo-beneficio que requiere la utilización de nuestras facultades "cognitivas", aquellas que se basan en la corteza prefrontal dorsolateral (Greene et al. 2004).

Esta explicación psicológica del consecuencialismo y la deontología da sentido a ciertos aspectos de sus fenomenologías asociadas. A menudo he observado que el consecuencialismo les resulta muy atractivo a los estudiantes, incluso tautológicamente cierto, cuando les es presentado en abstracto, pero su atractivo se ve fácilmente socavado mediante contraejemplos específicos (véase la discusión anterior que contrasta las motivaciones de las personas en el mundo real y las justificaciones abstractas para el castigo). Cuando un estudiante de ética

de primer año pregunta: "¿Pero no es obvio que uno debe hacer lo que produzca el mayor bien?", todo lo que tienes que hacer es exponer el caso del puente peatonal y habrás dejado claro tu punto. Con una sacudida emocional, cualquier atractivo "cognitivo" inicial que pudiesen tener los principios consecuencialistas son rápidamente neutralizados y el estudiante es de pronto un deontólogo recién convertido: "¿Por qué está mal empujar al hombre por el puente? ¡Porque tiene un derecho, una inviolabilidad fundada en la justicia que ni siquiera el bienestar de la sociedad en su conjunto puede invalidar!" Entonces, es hora de un nuevo contraejemplo: "¿Qué pasa si el carro se dirige a un detonador que hará explotar una bomba nuclear que matará a medio millón de personas?" De repente, el bienestar de la sociedad en su conjunto vuelve a parecer importante. La "cognición" contraataca con una lógica utilitarista más convincente y el estudiante quedó acertadamente desconcertado. Tal como lo ilustra esta dialéctica sencilla, la hipótesis de que la deontología se encuentra emocionalmente fundada explica la condición de tipo "¡NUNCA, excepto a veces!" de la ética deontológica basada en derechos. Una respuesta emocional de tipo alarma se presenta a sí misma como inflexible y absoluta, hasta el momento en que aparece una lógica emocional o "cognitiva" aún más convincente para anularla.

Esta hipótesis también le da sentido a ciertas anomalías deontológicas que sospecho que resultarán ser las "excepciones que prueban la regla". He argumentado que la deontología está impulsada por la emoción, pero sospecho que no siempre es así. Consideremos, por ejemplo, la infame afirmación de Kant de que sería incorrecto mentirle a un posible asesino para proteger a un amigo que se ha refugiado en la casa de uno (Kant 1983). Aquí, en una demostración dramática de verdadera integridad intelectual, Kant se apega a su teoría y rechaza la respuesta intuitiva ("muerde la bala", como dicen los filósofos). Pero lo interesante de esta parte de la ética kantiana es que representa una suerte de vergüenza para los kantianos contemporáneos y que, de hecho, están muy preocupados por explicar cómo en este caso Kant aplicó mal su propia teoría (Korsgaard 1996a). Presumiblemente, lo mismo ocurre con las opiniones de Kant sobre la moral sexual (Kant 1930, 169-171; Kant 1994). Los académicos modernos ya no son tan remilgados con la lujuria, la masturbación y la homosexualidad, por lo que a las opiniones anticuadas de Kant sobre estos temas deberían restársele importancia, lo que no es difícil, ya que para empezar sus argumentos nunca fueron terriblemente convincentes (ver epígrafe). Si quieres saber qué partes de Kant rechazarán los kantianos contemporáneos, sigue las emociones.

3. Implicaciones normativas

3.1 El "es" psicológico y el "deber" moral

Las hipótesis anteriormente planteadas sobre las respectivas bases psicológicas del consecuencialismo y la deontología podrían ciertamente estar equivocadas. Pero el que sean correctas o no, no puede determinarse desde el sillón. Más bien, es una cuestión empírica. Y aunque estas hipótesis siguen abiertas al desafío empírico, a partir de ahora voy a asumir que

son correctas para así explorar sus implicaciones filosóficas más amplias. Dado que la mayoría de los filósofos morales no consideran que sus puntos de vista dependan de los resultados de debates particulares en psicología experimental, este supuesto no debe considerarse excesivamente restrictivo.

De hecho, los filósofos morales tienden a mantenerse alejados de las controversias científicas siempre que les sea posible, asumiendo que los detalles científicos son en gran medida irrelevantes para su empresa: mientras que la ciencia se ocupa de lo que es, la moral se ocupa de lo que debería ser, y nunca se encontrarán (Hume 1978; Moore 1966). Contrariamente a esta sabiduría moral recibida, considero que la ciencia es importante para la ética, no porque uno pueda derivar verdades morales de verdades científicas, sino porque la información científica puede socavar los supuestos fácticos de los que depende implícitamente el pensamiento moral. El punto de contacto clave entre la filosofía moral y la psicología moral científica es la intuición moral. Los filósofos de la moral, desde Platón (1987) en adelante, se han basado en su sentido intuitivo del bien y el mal para guiarse en sus intentos de dar sentido a la moral. La relevancia de la ciencia, entonces, reside en que puede decirnos cómo funcionan y de dónde vienen nuestras intuiciones morales. Y una vez que entendamos un poco mejor nuestras intuiciones, podremos verlas de manera bastante diferente. Esto se aplica no solo a los moralistas que se basan explícitamente en las intuiciones morales (Ross 1930), sino también a los moralistas que desconocen hasta qué punto sus juicios morales están moldeados por la intuición.

En los últimos años, varios filósofos y científicos han cuestionado la fiabilidad de las intuiciones morales y han argumentado que comprender la psicología de la intuición moral posee implicaciones normativas (Baron 1994; Greene 2003; Horowitz 1998; Sinnott-Armstrong 2004; Unger 1996). Haré lo mismo, pero de una manera específica. Argumentaré que nuestra comprensión de la psicología moral, como se describió anteriormente, arroja dudas sobre la deontología como escuela de pensamiento moral normativo.

3.2 Racionalismo, racionalización y juicio deontológico

Tu amiga Alice tiene muchas citas y después de cada una te comenta lo sucedido. Cuando elogia a las personas que le agradan y se queja de las que no le gustan, cita una gran cantidad de factores. Esta es brillante. Esa es ensimismada. Esta tiene un gran sentido del humor. Esa es un fiasco. Y así. Pero luego notas algo: todas las personas que le gustan son excepcionalmente altas. Una inspección más cercana revela que, después de muchas citas durante varios años, no ha dado el visto bueno a nadie que mida menos de dos metros y medio, y no ha rechazado a nadie de esta altura (al conectar los datos de citas de Alice en un software de estadísticas se confirma que la altura es un predictor casi perfecto de sus preferencias). De repente, parece que el juicio de Alice no es lo que usted creía, y ciertamente no es lo que ella cree. Alice, por

supuesto, cree que sus juicios románticos se basan en una variedad de factores complejos. Pero, si confiamos en los números, básicamente tiene un fetiche con la altura y toda su charla sobre el ingenio, el encanto y la bondad es mera racionalización.

Lo que ilustra este ejemplo es que es posible detectar un racionalizador sin tener que separar el razonamiento del racionalizador. Tan solo se necesita hacer dos cosas. En primer lugar, se debe encontrar un factor que prediga los juicios del racionalizador. En segundo lugar, hay que demostrar que el factor que predice los juicios del racionalizador no está plausiblemente relacionado con los factores que, según el racionalizador, son la base de sus juicios. Usando esta estrategia, creo que uno puede argumentar bastante bien contra las versiones racionalistas de la deontología como la de Kant, es decir, aquellas según las cuales los juicios morales característicamente deontológicos se justifican en términos de teorías abstractas de derechos, deberes, etc. El caso en contra de tales teorías ya está implícito en el material empírico presentado anteriormente, pero vale la pena explicarlo.

La mayor parte de este capítulo estuvo dedicado a satisfacer el primero de los dos requisitos enumerados anteriormente, es decir, a identificar un factor en particular: la respuesta emocional que predice el juicio deontológico. A continuación, debemos considerar la naturaleza de la relación entre este factor predictivo y los factores que, según los deontólogos racionalistas, son la base de sus juicios. Por definición, un racionalista no puede sostener que alguna acción es correcta o incorrecta debido a las emociones que sentimos en respuesta a ella. Sin embargo, como una cuestión empírica de hecho (asumimos), hay una correspondencia notable entre lo que las teorías deontológicas racionalistas nos dicen que hagamos y lo que nuestras emociones nos dicen que hagamos. Así, a la luz de estos datos, hay una serie de coincidencias respecto de las cuales varios deontólogos racionalistas deben dar cuenta: por ejemplo, según Judith Jarvis Thomson (Thomson 1986; Thomson 1990) y Frances Kamm (Kamm 1993; Kamm 1996) (ambos cuentan como racionalistas para nuestros propósitos), existe una compleja y sumamente abstracta teoría de los derechos que explica por qué está bien sacrificar una vida por cinco en el caso del tranvía pero no en el caso del puente, y da la casualidad de que tenemos una fuerte respuesta emocional negativa al último caso pero no al primero. Asimismo, según Colin McGinn (McGinn 1999) y Frances Kamm (Kamm 1999), existe una teoría del deber que explica por qué tenemos la obligación de ayudar al niño que se está ahogando en el ejemplo de Singer pero no existe una obligación comparable de salvar a los niños hambrientos del otro lado del mundo, y da la casualidad de que tenemos fuertes respuestas emocionales hacia el primero pero no hacia los segundos. Según Kant (Kant 2002) y muchos otros teóricos del derecho (Lacey 1988; Ten 1987), existe una compleja y abstracta teoría del castigo que explica por qué deberíamos castigar a las personas independientemente de si se pueden obtener beneficios sociales al hacerlo y da la casualidad de que tenemos respuestas emocionales que nos inclinan a hacer exactamente eso. El imperativo categórico prohíbe la masturbación porque implica el uso de uno mismo como medio (Kant 1994), y da la casualidad de que el principal defensor del imperativo categórico encuentra la masturbación total y absolutamente repugnante (ver epígrafe). Y así sucesivamente.

Kant, como ciudadano de la Europa del siglo XVIII, tiene una explicación preparada para este tipo de coincidencias: Dios, en su sabiduría infinita, dotó a las personas de disposiciones emocionales diseñadas para animarlas a comportarse de acuerdo con la ley moral. Kant evitó invocar a Dios en sus argumentos filosóficos, pero es plausible pensar que su fe le impidió, junto con casi todos los demás de su época, sentirse desconcertado por el orden y la armonía del mundo natural, incluida su armonía con la ley moral. Además, a la luz de sus supuestos de fondo, no se puede culpar a Kant por tratar de racionalizar sus intuiciones morales. Sus intuiciones se derivan de su naturaleza humana ("la ley moral interna" (Kant 1993)) y, en última instancia, de Dios. Dios es un tipo inteligente, debió pensar Kant. No le daría a la gente intuiciones morales porque sí. En cambio, debemos tener las intuiciones que tenemos por buenas razones. Y así Kant se dispuso a descubrir esas razones, sino por la fuerza de la razón, sí por hazaña de la imaginación.

Los deontólogos racionalistas actuales, como ciudadanos del siglo XXI, no pueden depender de la idea de que Dios nos dio nuestras emociones morales para alentarnos a comportarnos de acuerdo con la verdad moral deontológica, racionalmente descubierta. En cambio, necesitan algún tipo de explicación naturalista respetable del hecho de que las conclusiones a las que llegan los deontólogos racionalistas, a diferencia de las alcanzadas por los consecuencialistas, parecen estar impulsadas por respuestas emocionales de tipo 'alarma'. A su vez, su explicación debe competir con la alternativa propuesta aquí, a saber, que las teorías deontológicas racionalistas son racionalizaciones de esas respuestas emocionales, la cual ya tiene cierta ventaja (a) por el hecho de que gran parte del comportamiento humano parece ser intuitivo (Bargh y Chartrand 1999) y (b) por la bien documentada tendencia de las personas a racionalizar su comportamiento intuitivo (Haidt 2001; Wilson 2002).

¿Qué tipo de explicación pueden dar los deontólogos racionalistas? Tendrán que sostener, en primer lugar, que la correspondencia entre el juicio deontológico y el compromiso emocional no es una coincidencia y, en segundo lugar, que nuestras emociones morales de alguna manera siguen la pista de la verdad moral deontológica que puede ser descubierta racionalmente. Sin embargo, siendo que son racionalistas, no pueden afirmar que nuestras respuestas emocionales son la base de la verdad moral. Por lo tanto, tendrán que explicar cómo una combinación de evolución biológica y cultural logró darnos disposiciones emocionales que corresponden a una verdad moral independiente, la cual se puede descubrir racionalmente y no se basa en la emoción.

Ya desde un principio hay otra desventaja a la que se enfrentan los encargados de esta tarea y consiste en el punto principal que deseo señalar en lo que sigue. Hay buenas razones para pensar que nuestras intuiciones morales distintivamente deontológicas (aquí, las que entran en conflicto con el consecuencialismo) reflejan la influencia de factores moralmente irrelevantes y, por lo tanto, es poco probable que sigan la verdad moral.

Tomemos, por ejemplo, los casos del tranvía y del puente. He argumentado que existe una distinción moral intuitiva entre estos dos casos fundada en el hecho de que mientras la

violación moral en el caso del puente es "cercana y personal", la violación moral en el caso del tranvía no lo es. Además, argumenté que respondemos de manera más emocional a las violaciones morales que son "cercanas y personales" porque ese es el tipo de violaciones que existían en el entorno en el que evolucionamos. En otras palabras, sostuve que tenemos una intuición característicamente deontológica con respecto al caso del puente debido a un rasgo contingente y no moral de nuestra historia evolutiva. A su vez, he argumentado que la misma hipótesis de lo "cercano y personal" da sentido a las desconcertantes intuiciones que rodean los ejemplos de Peter Singer y el efecto de la víctima identificable, lo que aumenta su poder explicativo.

El punto clave es que esta hipótesis es contraria a cualquier hipótesis según la cual nuestras intuiciones morales, en respuesta al tipo de casos mencionados, reflejan verdades morales profundas y capaces de ser descubiertas racionalmente. Por supuesto, la hipótesis que he propuesto podría estar equivocada. Ahora bien, ¿acaso los deontólogos racionalistas prefieren dar por hecho que está equivocada?, ¿tienen explicaciones positivas más plausibles que ofrecer en su lugar?

Una hipótesis similar puede explicar nuestras inclinaciones hacia el castigo retributivo. Los consecuencialistas sostienen que los castigos solo deben imponerse en la medida en que sea factible que produzcan buenas consecuencias (Bentham 1982). Los deontólogos como Kant (Kant 2002), junto con la mayoría de las personas (Baron et al. 1993; Baron y Ritov 1993), son retributivistas. Juzgan a favor de castigar a los malhechores como un fin en sí mismo, incluso cuando resulta poco probable que el hacerlo promueva buenas consecuencias en el futuro. ¿Es esto una percepción moral de su parte o simplemente un subproducto de nuestra psicología evolucionada? La evidencia disponible sugiere lo último.

Según lo discutido anteriormente, parece que las emociones que nos impulsan a castigar a los malhechores evolucionaron como un mecanismo eficiente para estabilizar la cooperación, tanto entre individuos (Trivers 1971) como dentro de grupos más grandes (Bowles y Gintis 2004; Boyd et al. 2003; Fehr y Gachter 2002). En otras palabras, según estos modelos, estamos dispuestos a castigar debido a las consecuencias biológicas de esta disposición. Además, la selección natural, al proporcionarnos esta disposición, "eligió", por así decirlo. Es decir, por un lado, la naturaleza podría habernos dado una disposición para castigar otorgándonos, en primer lugar, un deseo innato de asegurar los beneficios de la cooperación futura y, en segundo lugar, algunos medios para reconocer que castigar a los que no cooperan es a menudo una buena manera de lograr este fin. En otras palabras, la naturaleza podría habernos convertido en consecuencialistas del castigo. Por otro lado, la otra opción era que la naturaleza nos dé un deseo directo de castigar a los no cooperadores como un fin en sí mismo, incluso si castigar en algunos casos no genera ningún bien (biológico). Como se señaló anteriormente, la naturaleza enfrenta este tipo de elección cada vez que genera una adaptación conductual y, en casi todos los casos, la naturaleza adopta el enfoque más directo: hablando en términos psicológicos, deseamos cosas como comida, sexo y un lugar cómodo para descansar porque son agradables (y porque su ausencia es desagradable) y no porque creamos que mejorarán nuestra

aptitud biológica. La disposición al castigo no parece ser una excepción a este patrón general. Psicológicamente hablando, castigamos principalmente porque encontramos que el castigo es satisfactorio (de Quervain et al. 2004) y porque consideramos que las transgresiones impunes resultan claramente insatisfactorias (Carlsmith et al. 2002; Kahneman et al. 1998; Sanfey et al. 2003).).

En otras palabras, las emociones que nos impulsan a castigar son instrumentos biológicos contundentes. Evolucionaron porque nos impulsan a castigar de maneras que conducen a buenas consecuencias (en términos biológicos). Pero, como subproducto de su diseño simple y eficiente, también nos llevan a castigar en situaciones en las que no se pueden esperar consecuencias (en términos biológicos) buenas. Por lo tanto, tal parece que, como cuestión evolutiva de hecho, tenemos una preferencia por la retribución no porque los malhechores realmente merezcan ser castigados independientemente de los costos y beneficios, sino porque las disposiciones retributivas son una forma eficiente de inducir un comportamiento que favorece que los individuos vivan en grupos sociales para así difundir más eficazmente sus genes.

Por supuesto, es posible que aquí haya una coincidencia. Quizás sea parte de la verdad moral, aquella que es posible descubrir racionalmente, según la cual las personas merecen ser castigadas como un fin en sí mismo. Al mismo tiempo, podría suceder que la selección natural, al idear un medio eficaz para promover consecuencias biológicamente ventajosas, nos proporcione disposiciones de base emocional que nos lleven a esta conclusión. Pero esto parece poco probable. Más bien, parece que las teorías retributivistas del castigo son solo racionalizaciones de nuestros sentimientos retributivistas y, a su vez, que estos sentimientos solo existen debido a las restricciones que se imponen a la selección natural, irrelevantes a nivel moral, al momento de diseñar criaturas que buscan mejorar su aptitud y poseen comportamientos acordes a ello. En otras palabras, la historia natural de nuestras disposiciones retributivistas hace que sea poco probable que reflejen algún tipo de verdad moral profunda.

Preciso enfatizar que no estoy afirmando que las teorías consecuencialistas del castigo sean correctas porque la tendencia a castigar evolucionó para producir buenas consecuencias. Suponer que estas "buenas consecuencias" solo necesiten ser buenas desde un punto de vista biológico y asumir que nuestros fines deben coincidir con los fines de la selección natural sería caer en la falacia naturalista en su forma original (Moore 1966). Al mismo tiempo, quiero dejar en claro que no estoy afirmando que cualquier tendencia que tengamos como derivada de nuestra evolución sea automáticamente incorrecta o equivocada. No diría, por ejemplo, que está mal amar a hijos adoptados (que no comparten los genes de uno) o usar métodos anticonceptivos simplemente porque son comportamientos que frustran las "intenciones" de la naturaleza. Lo que afirmo, en este punto, es simplemente que es poco probable que las inclinaciones que evolucionaron como subproductos evolutivos correspondan a alguna verdad moral independiente que se pueda descubrir racionalmente. En cambio, es más parsimonioso

suponer que cuando nos sentimos atraídos por las teorías retributivistas del castigo es porque estamos gravitando hacia nuestras inclinaciones emocionales evolucionadas y no hacia alguna verdad moral independiente[16].

Lo que la ciencia del cambio de milenio nos está diciendo es que el juicio moral humano no es una empresa racional impoluta, sino que nuestros juicios morales están impulsados por una mezcolanza de disposiciones emocionales que, a su vez, fueron moldeadas por una mezcolanza de fuerzas evolutivas, tanto biológicas como culturales. Debido a esto, es extremadamente improbable que exista alguna teoría moral normativa racionalmente coherente que pueda hacer encajar nuestras intuiciones morales. Además, es casi seguro que cualquiera que afirme tener tal teoría, o incluso una parte, en realidad no la tenga. En cambio, es probable que lo que esa persona tenga sea una racionalización moral.

Tal parece, entonces, que de alguna manera hemos cruzado la infame división entre el "es" y el "debe"[17]. ¿Cómo sucedió esto? ¿No nos habían advertido ya Hume (Hume 1978) y Moore (Moore 1966) sobre la pretensión de derivar un "debe" de un "es"? ¿Cómo pasamos de las teorías científicas descriptivas sobre la psicología moral al escepticismo respecto de toda una clase de teorías morales normativas? La respuesta es que no hemos intentado, como anticiparon Hume y Moore, derivar un "debe" de un "es". Es decir, nuestro método ha sido inductivo más que deductivo. Sobre la base de la evidencia disponible, hemos inferido que el fenómeno de la filosofía deontológica racionalista se explica mejor como una racionalización de intuiciones emocionales evolucionadas (Harman 1977).

3.3 Pasando por alto el punto deontológico

Sospecho que los deontólogos racionalistas permanecerán impasibles ante los argumentos aquí presentados. En cambio, conjeturo, insistirán en que simplemente he entendido mal los desarrollos de Kant y deontólogos afines. La deontología, dirán, no se trata de esta o aquella intuición. Tampoco se define por sus diferencias normativas con el consecuencialismo. Más bien, la deontología consiste en tomarse en serio la humanidad. Se trata, por sobre todo, de respetar a las personas. Es decir, de tratar a los demás como criaturas racionales en lugar de como meros objetos, de actuar por razones que los seres racionales comparten. Y así sucesivamente (Korsgaard 1996a; Korsgaard 1996b).

Sin duda, es de esta manera como muchos deontólogos ven la deontología. No obstante, esta perspectiva puertas adentro, según he sugerido, puede ser engañosa. El problema de di-

[16] Es decir, una verdad independiente de los detalles de la psicología moral humana y de los acontecimientos naturales que la moldearon.

[17] La mayoría está de acuerdo en que la división entre el "es" y el "debe" puede ser cruzada cuando el "es" equivale a una restricción sobre lo que se puede hacer y, a fortiori, es una restricción sobre lo que "debe" hacerse. Por ejemplo, si es el caso de que estás muerto, entonces no es el caso de que debas votar. Sin embargo, el paso del "es" al "debe" que se analiza más adelante es más sustantivo y, en consecuencia, más controvertido.

cho enfoque, más específicamente, es que define la deontología en términos de valores que no son distintivamente deontológicos, aunque desde el interior puedan parecerlo así. Consideremos la siguiente analogía con la religión. Cuando uno le pide a una persona religiosa que explique la esencia de su religión, a menudo obtiene una respuesta como esta: "Se trata de amor, realmente. Se trata de cuidar a los demás, de mirar más allá de uno mismo. De estar en comunidad, de formar parte de algo más grande que uno mismo". Este tipo de respuesta captura con precisión la fenomenología religiosa de muchas personas pero, no obstante, es inadecuada para distinguir la religión de otras cosas. Esto se debe a que muchas personas no religiosas, si no la mayoría, aspiran a amar profundamente, a cuidar de otros, a evitar el ensimismamiento, a tener un sentido de comunidad y a estar conectadas con cosas más grandes que ellas mismas. En otras palabras, los humanistas y ateos seculares pueden estar de acuerdo con la mayor parte de lo que muchas personas religiosas consideran que define a la religión. Por el contrario, desde el punto de vista de un humanista secular, lo distintivo de la religión es su compromiso con la existencia de entidades sobrenaturales, así como con instituciones y doctrinas religiosas formales. Y tienen razón. Estas son las cuestiones que realmente distinguen las prácticas religiosas de las no religiosas, aunque pueden parecer secundarias para muchas personas que operan desde un punto de vista religioso.

De la misma manera, creo que la mayoría de las caracterizaciones deontológicas/kantianas estándar no logran distinguir la deontología de otros enfoques de la ética (véase también Kagan [1997, 70-78] sobre la dificultad de definir la deontología). Considero que los consecuencialistas, tanto como cualquiera, tienen respeto por las personas, están en contra de tratarlas como meros objetos, desean actuar por razones que las criaturas racionales pueden compartir, etc. Un consecuencialista respeta a los demás y se abstiene de tratarlos como meros objetos, incluyendo el bienestar de cada persona en el proceso de toma de decisiones. A su vez, un consecuencialista intenta actuar de acuerdo con razones que las criaturas racionales pueden compartir, es decir, cuando actúan en conformidad con principios que dan igual peso a los intereses de todos y que, por tanto, son imparciales. Esto no quiere decir que los consecuencialistas y los deontólogos no difieran entre sí. De hecho, lo hacen. A lo que me refiero es a que las diferencias reales pueden no ser lo que los deontólogos a menudo creen que son.

Entonces, ¿qué distingue a la deontología de otros tipos de pensamiento moral? Una buena estrategia para responder a esta pregunta es comenzar con desacuerdos concretos entre los deontólogos y otros puntos de vista (como el de los consecuencialistas), para luego trabajar hacia atrás en busca de principios más profundos. Esto es lo que he intentado hacer con los casos del tranvía y del puente, y en otros casos en los que los deontólogos y los consecuencialistas no están de acuerdo. Si usted le pregunta a alguien de mentalidad deontológica por qué no está bien empujar a una persona frente a un tranvía a alta velocidad para salvar a otras cinco, obtendrá respuestas característicamente deontológicas. Algunos serán tautológicos: "¡Porque es un asesinato!". Otros serán más sofisticados: "El fin no justifica los medios", "Tienes que respetar los derechos de las personas". Pero, como sabemos, estas respuestas en realidad no explican nada, porque si a los mismos individuos (en diferentes ocasiones) le presentas el

caso del tranvía o el caso del bucle (ver arriba), harán el juicio opuesto, incluso aunque su explicación inicial sobre el caso del puente se aplique igualmente bien a uno o ambos casos. Hablar sobre los derechos, el respeto por las personas y las razones que podemos compartir son intentos naturales de explicar, en términos "cognitivos", lo que sentimos cuando nos encontramos con intuiciones emocionalmente impulsadas que entran en conflicto con el frío cálculo del consecuencialismo. Aunque estas explicaciones son inevitablemente incompletas, parece haber "algo profundamente correcto" en ellas porque, justamente, dan voz a poderosas emociones morales. Pero, como ocurre con los relatos de muchas personas religiosas sobre lo que es esencial a la religión, en realidad no explican qué es lo distintivo de la filosofía en cuestión.

En resumen, si acaso parece que simplemente he entendido mal en qué consisten Kant y la deontología, es porque estoy proponiendo una hipótesis alternativa a la comprensión estándar kantiana/deontológica de Kant y la deontología. Estoy planteando una hipótesis empírica sobre la esencia psicológica oculta de la deontología, y esta no puede ser descartada a priori por la misma razón por la que los isleños tropicales no pueden saber a priori si el hielo es una forma de agua.

3.4 Psicología moral evolutiva y moral antropocéntrica

Anteriormente expuse un caso en contra de la deontología racionalista, a saber, la idea de que nuestras intuiciones morales deontológicas pueden justificarse mediante teorías abstractas de derechos, deberes, etc. Sin embargo, existen formas de deontología más modestas. En lugar de defender nuestras intuiciones morales bajo el supuesto de que pueden ser justificadas por una teoría racional, podríamos simplemente apoyarlas porque son nuestras. Es decir, uno podría adoptar un enfoque antropocéntrico de la moral (ver Haidt y Bjorklund, en este volumen), renunciando al sueño de la Ilustración de derivar las verdades morales de primeros principios y estableciendo, en cambio, una moral que es contingentemente humana.

Esta es la dirección en la que, en las últimas décadas, se ha movido la filosofía moral. La ética de la virtud define la bondad moral en términos del carácter humano (Crisp y Slote 1997; Hursthouse 1999). Los "teóricos de la sensibilidad", de ideas afines, consideran que ser moral es una cuestión de tener el tipo correcto de sensibilidad distintivamente humana (McDowell 1988; Wiggins 1987). Los especialistas en ética con una inclinación más metafísica hablan de propiedades morales que son "dependientes de la respuesta" (Johnston 1995), sentimientos morales que corresponden a propiedades morales "cuasi-reales" (Blackburn 1993) y propiedades morales que son "grupos homeostáticos" de propiedades naturales (Boyd 1988). Incluso dentro de la tradición kantiana muchos enfatizan la "construcción" de principios morales que, en lugar de ser verdaderos, son "razonables para nosotros" (Rawls 1995) o, alternativamente, las demandas normativas que se derivan de nuestras "identidades prácticas" distintivamente humanas (Korsgaard 1996b).

En resumen, la filosofía moral reciente es decididamente antropocéntrica en el sentido de que muy pocos filósofos están desafiando activamente las intuiciones morales de alguien. Reconocen que nuestras virtudes morales, sensibilidades, identidades, etc. pueden cambiar con el tiempo pero, al menos la mayoría, no están activamente tratando de cambiarlas.

El argumento presentado anteriormente crea problemas para aquellos en busca de teorías racionalistas que puedan explicar y justificar sus intuiciones morales deontológicas impulsadas emocionalmente. Ahora bien, los deontólogos racionalistas pueden no ser los únicos que deberían pensarlo dos veces. Los argumentos anteriormente presentados arrojan dudas sobre las intuiciones morales en cuestión, independientemente de si se desea justificarlas en términos teóricos abstractos. Esto se debe, una vez más, a que estas intuiciones parecen haber sido moldeadas por factores moralmente irrelevantes relacionados con las limitaciones y circunstancias de nuestra historia evolutiva. De modo que se trata de un problema para cualquiera que esté inclinado a apoyar estas intuiciones y, vale resaltar, "cualquiera" incluye a casi todos.

Me he referido a estas intuiciones y los juicios que respaldan como "deontológicos", pero tal vez sería más exacto llamarlos no consecuencialistas (Baron 1994). Después de todo, no es necesario ser un deontólogo militante para pensar que está bien comer en restaurantes cuando la gente del mundo se muere de hambre, que es intrínsecamente bueno que los delincuentes sufran por sus delitos y que sería un error empujar al tipo del puente. Estos juicios son perfectamente de sentido común y parece que las únicas personas que se inclinan a cuestionarlos son los consecuencialistas militantes.

¿Significa esto que todos los no consecuencialistas necesitan repensar al menos algunos de sus compromisos morales? Humildemente sugiero que la respuesta es sí. Consideremos, una vez más, el argumento de Peter Singer sobre las obligaciones morales que conlleva la opulencia. Supongamos, una vez más, que los hechos evolutivos y psicológicos son exactamente como he desarrollado. Es decir, supongamos que la única razón por la que decimos que, por un lado, está mal abandonar al niño que se ahoga y que, por otro lado, está bien ignorar las necesidades de los niños hambrientos en el extranjero, es porque el primero empuja nuestros botones emocionales y el segundo no. Supongamos, además, que la única razón por la que los niños que están lejos de nosotros no logran presionar nuestros botones emocionales es porque evolucionamos en un entorno en el que era imposible interactuar con individuos lejanos. ¿Podríamos entonces defender nuestras intuiciones de sentido común? ¿Podemos, con la conciencia tranquila, decir: "vivo una vida de lujos mientras ignoro las necesidades desesperadas de las personas que están lejos porque, por un accidente de la evolución humana, soy emocionalmente insensible a su difícil situación y, sin embargo, mi fracaso en aliviar su sufrimiento, cuando fácilmente podría hacer algo diferente al respecto, está perfectamente justificado"? No sé ustedes, pero encuentro incómoda esta combinación de afirmaciones. Esto no quiere decir, por supuesto, que me sienta cómodo con la idea de renunciar a la mayoría de mis posesiones y privilegios mundanos para ayudar a extraños. Después de todo, soy solo un ser humano. Pero, al menos para mí, comprender la fuente de mis intuiciones morales inclina la balanza, en este caso y en otros, en una dirección más singeriana y consecuencialista.

Así, como resultado de comprender los hechos psicológicos, soy menos complaciente con mi tendencia demasiado humana a ignorar el sufrimiento distante. Del mismo modo, cuando entiendo las raíces de mis impulsos retributivos, es menos probable que les conceda autoridad moral. Lo mismo sucede con cualquier obsesión que pueda tener sobre el comportamiento sexual desviado pero inofensivo.

Sin embargo, tomar estos argumentos en serio amenaza con ponernos en una segunda pendiente resbaladiza (además de la que conduce a la indigencia altruista): ¿Cuán lejos puede llegar la desmitificación empírica de la naturaleza moral humana? Si la ciencia me dice que amo a mis hijos más que a otros niños solo porque comparten mis genes (Hamilton 1964), ¿debería sentirme incómodo por quererlos más? Si la ciencia me dice que soy amable con otras personas solo porque la disposición a ser amable ayudó a mis antepasados a difundir sus genes (Trivers 1971), ¿debería dejar de ser amable con la gente? Si me preocupo por mí mismo solo porque estoy biológicamente programado para reproducir mis genes, ¿debería dejar de preocuparme por mí mismo? Parece ser que alguien que no está dispuesto a actuar sobre las tendencias humanas que tienen causas evolutivas amorales, en última instancia, no está dispuesto a ser humano. ¿Dónde se traza la línea divisoria entre corregir la miopía de la naturaleza moral humana y borrarla por completo?

Creo que esta es una de las cuestiones morales más fundamentales que enfrentamos en una era de creciente autoconocimiento científico y no intentaré abordarla aquí. En otro lugar sostengo que los principios consecuencialistas, aunque no son ciertos, proporcionan el mejor estándar disponible para la toma de decisiones públicas y para determinar qué aspectos de la naturaleza humana es razonable intentar cambiar y cuáles sería prudente dejar en paz (Greene 2002; Greene y Cohen 2004).

Agradecimientos

Muchas gracias a Walter Sinnott-Armstrong, Jonathan Haidt, Shaun Nichols y Andrea Heberlein por sus útiles comentarios sobre este capítulo.

Referencias bibliográficas

Adolphs R. (2002). Neural systems for recognizing emotion. *Current opinion in neurobiology*, *12*(2), 169-177. https://doi.org/10.1016/s0959-4388(02)00301-x

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in cognitive sciences*, *4*(7), 267-278. https://doi.org/10.1016/s1364-6613(00)01501-1

Allman, J., Hakeem, A., Watson, K. (2002). Two phylogenetic specializations in the human brain. *Neuroscientist*, *8*(4), 335-346.

Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., Catroppa, C. (2001). Development of executive functions through late childhood and adolescence in an Australian sample. *Developmental neuropsychology*, *20*(1), 385-406. https://doi.org/10.1207/S15326942DN2001_5

Aquinas, T. (1988). Of killing. In Baumgarth, W. P., Regan, R. J. (Eds.), *On law, morality, and politics*, pp. 226-227. Indianapolis/Cambridge: Hackett Publishing Co.

Bargh, J. A., Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*, 462-479.

Baron, J. (1994). Nonconsequentialist decisions. *Behavoral and Brain Sciences*, *17*, 1-42.

Baron, J., Gowda, R., Kunreuther, H. (1993). Attitudes toward managing hazardous waste: What should be cleaned up and who should pay for it? *Risk Analysis*, *13*(2), 183-192.

Baron, J., Ritov, I. (1993). Intuitions about penalties and compensation in the context of tort law. *Journal of Risk and Uncertainty*, *7*(1), 17-33. https://doi.org/10.1007/BF01065312

Bentham, J. (1982). *An introduction to the principles of morals and legislation.* Londres: Methuen.

Blackburn, S. (1993). *Essays in quasi-realism*. New York: Oxford University Press.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624-652.

Bowles, S., Gintis, H. (2004). The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theoretical population biology*, *65*(1), 17-28. https://doi.org/10.1016/j.tpb.2003.07.001

Boyd, R., Gintis, H., Bowles, S., Richerson, P. J. (2003). The evolution of altruistic punishment. *PNAS*, *100*(6), 3531-3535.

Boyd, R. N. (1988). How to be a moral realist. In, Sayre-McCord, G. (Ed.) *Essays on moral realism*, pp. 181-228. Ithica, NY: Cornell University Press.

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of personality and social psychology*, *83*(2), 284-299. https://doi.org/10.1037/0022-3514.83.2.284

Chalmers, D. J. (1996). *The conscious mind. In search of a fundamental theory*. New York: Oxford University Press.

Crisp, R., Slote, M., (Eds.) (1997). *Virtue ethics*. New York: Oxford University Press.

de Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, *305*, 1254-1258.

Dutton, D. G., Aron, A. P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of personality and social psychology*, *30*(4), 510-517. https://doi.org/10.1037/h0037031

Fehr, E., y Gachter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137-140.

Fehr, E., & Rockenbach, B. (2004). Human altruism: economic, neural, and evolutionary perspectives. *Current opinion in neurobiology*, *14*(6), 784-790. https://doi.org/10.1016/j.conb.2004.10.007

Fischer, J. M., Ravizza, M., (Eds.) (1992). *Ethics: Problems and principles*. Fort Worth: Harcourt Brace Jovanovich College Publishers.

Foot, P. (1978). The problem of abortion and the doctrine of double effect. En *Virtues and vices*, pp. 19-32. Oxford: Blackwell.

Gazzaniga, M. S., Le Doux, J. E. (1978). *The integrated mind*. New York: Plenum.

Gintis, H. (2000). Strong reciprocity and human sociality. *Jornal of Theoretical Biology*, *206*(2), 169-179. https://doi.org/10.1006/jtbi.2000.2111

Goel, V., Dolan, R. J. (2004). Differential involvement of left prefrontal cortex in inductive and deductive reasoning. *Cognition*, *93*(3), B109-B121. https://doi.org/10.1016/j.cognition.2004.03.001

Greene, J. (2002) *The terrible, horrible, no good, very bad truth about morality and what to do about it*. Doctoral Thesis, Princeton University.

Greene, J. (2003). From neural 'is' to moral 'ought': What are the moral implications of neuroscientific moral psychology? *Nature Reviews Neuroscience*, *4*, 846-849.

Greene, J., Cohen, J. (2004). For the law, neuroscience changes nothing and everything. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *359*(1451), 1775-1785. https://doi.org/10.1098/rstb.2004.1546

Greene, J., Haidt, J. (2002). How (and where) does moral judgment work?. *Trends in cognitive sciences*, *6*(12), 517-523. https://doi.org/10.1016/s1364-6613(02)02011-9

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44, 389-400.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., y Cohen, J. D. (2001). An fmri investigation of emotional engagement in moral judgment. *Science* 293, 2105-2108.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814-834.

Haidt, J., Bjorklund, F., Murphy, S. (2000). Moral dumbfounding: When intuition finds no reason. (unpublished manuscript, university of virginia).

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of personality and social psychology*, *65*(4), 613-628. https://doi.org/10.1037//0022-3514.65.4.613

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology*, 7, 1-52.

Harman, G. (1977). *The nature of morality*. New York: Oxford University Press.

Horowitz, T. (1998). Philosophical intuitions and psychological theory. In DePaul, M., Ramsey, W. (Eds.) *Rethinking intuition*, pp. 57-72. Lanham: Rowman y Littlefield Publishers.

Hume, D. (1978). A treatise of human nature. In Selby-Bigge, L.A., Nidditch, P. H. (Eds.) *A treatise of human nature*, pp. xix, 743. Oxford: Oxford University Press.

Hursthouse, R. (1999). *On virtue ethics*. New York: Oxford University Press.

Johnston, M. (1995). Dispositional theories of value. In Smith, M. (Ed.) *Meta-ethics*, pp. 113-137. Aldershot: Dartmouth Publishing Group.

Kagan, S. (1989). *The limits of morality*. New York: Oxford University Press.

Kagan, S. (1997). *Normative ethics*. Boulder: Westview Press.

Kahneman, D., Schkade, D., Sunstein, C. R. (1998). Shared outrage and erratic rewards: The psychology of punitive damages. *Journal of Risk and Uncertainty*, *16*, 49-86.

Kamm, F. M. (1993). *Morality, mortality*. Vol. 1: Death and whom to save from it. New York: Oxford University Press.

Kamm, F. M. (1996). *Morality, mortality*. vol. II: Rights, duties, and status. New York, Oxford University Press.

Kamm, F. M. (1999). Faminine ethics: The problem of distance in morality and singer's ethical theory. In Jamieson, D. (Ed.) *Singer and his critics*, 162-208. Oxford: Blackwell.

Kant, I. (1930). *Lectures on ethics*. Indianapolis/Cambridge: Hackett.

Kant, I. (1959). *Foundation of the metaphysics of morals*. Indianapolis: Bobbs-Merrill.

Kant, I. (1983). On a supposed right to lie because of philanthropic concerns. In Abbott, T. K. (trans.), *Kant's Critique of Practical Reason and Other Works on the Theory of Ethics,* pp. 162-166. Indianapolis/Cambridge: Hackett.

Kant, I. (1993). *Critique of practical reason*. Upper Saddle River: Prentice-Hall.

Kant, I. (1994). The metaphysics of morals. In Ethical philosophy (Indianapolis, IA, Hackett).

Kant, I. (2002). *The philosophy of law: An exposition of the fundamental principles of jurisprudence as the science of right*. Union: Lawbook Exchange.

Koechlin, E., Basso, G., Pietrini, P., Panzer, S., Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature*, *399*, 148-151.

Koechlin, E., Ody, C., Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, *302*, 1181-1185.

Kohlberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. InMischel, T. (Ed.) *Cognitive development and epistemology*, pp. 151-235. New York: Academic Press.

Korsgaard, C. M. (1996a). *Creating the kingdom of ends*. New York: Cambridge University Press.

Korsgaard, C. M. (1996b). *The sources of normativity*. New York: Cambridge University Press.

Kripke, S. A. (1980). *Naming and necessity*. Cambridge: Harvard University Press.

Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.

Lacey, N. (1988). *State punishment: Political principles and community values*. Londres; New York: Routledge & Kegan Paul.

Lieberman, D., Tooby, J., Cosmides, L. (2003). Does morality have a biological basis? An empirical test of the factors governing moral sentiments relating to incest. *Proceedings. Biological sciences*, *270*(1517), 819-826. https://doi.org/10.1098/rspb.2002.2290

Maddock R. J. (1999). The retrosplenial cortex and emotion: new insights from functional neuroimaging of the human brain. *Trends in neurosciences*, *22*(7), 310-316. https://doi.org/10.1016/s0166-2236(98)01374-5

Maier, N. R. F. (1931). Reasoning in humans. II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology, 12*(2), 181-194. https://doi.org/10.1037/h0071361

McClure, S. M., Laibson, D. I., Loewenstein, G., Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science, 306*, 503-507.

McDowell, J. (1988). Values and secondary qualities. In Sayre-McCord, G. (Ed.), *Essays on moral realism*, pp. 168-180. Ithaca: Cornell University Press.

McGinn, C. (1999). Out duties to animals and the poor. In Jamieson, D. (Ed.) *Singer and his critics*, pp. 150-161. Oxford: Blackwell.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, *24*, 167-202. https://doi.org/10.1146/annurev.neuro.24.1.167

Moore, G. E. (1966). *Principia ethica*. Cambridge: Cambridge University Press.

Nietzsche, F. (1974). *The gay science*. New York: Random House.

Nisbett, R. E., Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231-259.

Paus, T., Zijdenbos, A., Worsley, K., Collins, D. L., Blumenthal, J., Giedd, J. N., Rapoport, J. L., Evans, A. C. (1999). Structural maturation of neural pathways in children and adolescents: In vivo study. *Science*, *283*, 1908-1911.

Petrinovich, L., O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology*, *17*, 145-171.

Petrinovich, L., O'Neill, P., Jorgensen, M. (1993). An empirical study of moral intuitions: Toward an evolutionary ethics. *Journal of Personality and Social Psychology*, *64*, 467-478.

Phan, K. L., Wager, T., Taylor, S. F., Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage*, *16*(2), 331-348. https://doi.org/10.1006/nimg.2002.1087

Pinker, S. (2002). *The blank slate: The modern denial of human nature*. New York: Viking.

Plato (1987). *The republic*. Nueva York: Penguin Classics.

Putnam, H. (1975). The meaning of 'meaning'. In Gunderson, K. (Ed.) *Language, mind, and knowledge*, pp. 131-193. Minneapolis: University of Minnesota Press.

Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nature reviews. Neuroscience*, *5*(3), 184-194. https://doi.org/10.1038/nrn1343

Rawls, J. (1971). *A theory of justice*. Cambridge: Harvard University Press.

Rawls, J. (1995). Construction and objectivity. In Smith, M. (Ed.) *Meta-ethics*, pp. 231-252. Aldershot: Dartmouth Publishing Group.

Richerson, P. J., y Boyd, R. (2005). Not by genes alone: How culture transformed human evolution (Chicago, University of Chicago Press).

Ross, W. D. (1930). *The right and the good*. Oxford: Oxford University Press.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science*, *300*(5626), 1755-1758. https://doi.org/10.1126/science.1082976

Saxe, R., Carey, S., Kanwisher, N. (2004). Understanding other minds: linking developmental psychology and functional neuroimaging. *Annual review of psychology*, *55*, 87-124. https://doi.org/10.1146/annurev.psych.55.090902.142044

Schelling, T. C. (1968). The life you save may be your own. Chase, S. B. (ed.) In *Problems in public expenditure analysis*. Washington, D.C.: Brookings Institute.

Schnall, S., Haidt, J., Clore, G. (2004). Irrelevant disgust makes moral judgment more severe, for those who listen to their bodies.

Singer, P. (1972). Famine, affluence and morality. *Philosophy and Public Affairs*, *1*, 229-243.

Sinnott-Armstrong, W. (2006) Moral Intuitionism Meets Empirical Psychology. En Horgan, T., Timmons, M. (eds), *Metaethics after Moore,* 339-366. Oxford: Oxford University Press.

Small, D. A., Loewenstein, G. (2003). Helping a victim or helping the victim. *Journal of Risk and Uncertainty*, *26*, 5-16. https://doi.org/10.1023/A:1022299422219

Small, D. A., Lowenstein, G. (in press). The devil you know: the effects of identifiability on punitiveness. *Journal of Behavioral Decision Making*.

Smith, A. (1976). *The theory of moral sentiments*. Oxford: Oxford University Press.

Sober, E., Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior* Cambridge: Harvard University Press.

Steinberg, L., & Scott, E. S. (2003). Less guilty by reason of adolescence: developmental immaturity, diminished responsibility, and the juvenile death penalty. *The American psychologist*, *58*(12), 1009-1018. https://doi.org/10.1037/0003-066X.58.12.1009

Stuss, D. T., Alexander, M. P., Lieberman, A., Levine, H. (1978). An extraordinary form of confabulation. *Neurology*, *28*, 1166-1172.

Ten, C. L. (1987). *Crime, guilt, and punishment*. Oxford: Clarendon Press.

Thomson, J. J. (1986). *Rights, restitution, and risk: Essays, in moral theory*. Cambridge: Harvard University Press.

Thomson, J. J. (1990). *The realm of rights*. Cambridge: Harvard University Press.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46*, 35-57.

Turiel, E., Killen, M., Helwig, C. C. (1987). Morality: Its structure, function, and vagaries. In Kagan, J., Lamb, S. (Eds), *The emergence of morality in young children*. pp. 155-243. Chicago: University of Chicago Press.

Unger, P. K. (1996). *Living high and letting die: Our illusion of innocence*. New York: Oxford University Press.

Variety (1989). Tv reviews--network: Everybody's baby.

Weiner, B., Graham, S., Reyna, C. (1997). An attributional examination of retributive versus utilitarian philosophies of punishment. *Social Justice Research*, *10*, 431-452.

Wheatley, T., Haidt, J. (2005). Hypnotically induced disgust makes moral judgments more severe. *Psychological Science*, *16*(10), 780-784.

Wiggins, D. (1987*). Needs, values, and truth: Essays in the philosophy of value*. Oxford: Blackwell.

Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge: Harvard University Press.

Wrangham, R., Peterson, D. (1996). *Demonic males: Apes and the origins of human violence*. Boston: Houghton Mifflin.

Wright, R. (1994). *The moral animal: Evolutionary psychology and everyday life*. New York: Pantheon.