

**LA NEUROPSICOLOGÍA DEL JUICIO MORAL.  
SOBRE LAS CAUSAS DE RESPUESTAS CONTRAINTUITIVAS  
A LOS DILEMAS MORALES**

**The neuropsychology of moral judgment. About the causes  
of counter-intuitive responses to moral dilemmas**

*Alejandro Rosas  
María Andrea Arciniegas  
Esteban Caviedes  
María Alejandra Arciniegas*  
Universidad Nacional de Colombia

**Resumen**

*Una literatura creciente en neuropsicología estudia el juicio moral aplicando dilemas morales a personas con daños neuronales o con rasgos de personalidad antisocial; y parece confirmar una tendencia al juicio utilitarista contra-intuitivo, es decir, estas personas aprueban el sacrificio de un inocente para salvar varias vidas. Argumentamos que las evidencias encontradas no respaldan la hipótesis de dos módulos morales (Greene et al. 2004), sino más bien una disfunción en la capacidad empática o en el procesamiento de emociones, que apoyaría una psicología moral de corte Humeano. Abogamos, además, por ampliar el explanandum: del juicio utilitarista contra-intuitivo al juicio moral contra-intuitivo sin más. Esto permitiría investigar experimentalmente un amplio rango de variables afectadas por disfunciones en la capacidad para la empatía.*

Palabras Clave: Dilemas morales, juicio moral, daños neuronales, empatía, utilitarismo, deontología

---

**Recibido: agosto 18 de 2013    Aprobado: octubre 23 de 2013**

# **THE NEUROPSYCHOLOGY OF MORAL JUDGMENT. ABOUT THE CAUSES OF COUNTER-INTUITIVE RESPONSES TO MORAL DILEMMAS**

## *Abstract*

*A growing literature in neuropsychology studies moral judgment with moral dilemmas, testing subjects with neural damage or with anti-social personality traits. It seems to confirm a tendency to counterintuitive utilitarian moral judgment, i.e., these subjects abnormally approve sacrificing an innocent person to save several lives. We argue that empirical evidence does not support the hypothesis of two moral subsystems in the brain, but rather a dysfunction in empathic capacity and emotional processing, that provides support to a Humean moral psychology. Additionally, we advocate extending the explanandum of this research program from counterintuitive utilitarian judgment to counterintuitive moral judgment. This would allow to experimentally investigate a broader range of variables affected by dysfunction in empathic capacities.*

Keywords: Moral dilemmas, moral judgment, neural damage, empathy, utilitarianism, deontology

**Alejandro Rosas.** PhD Universidad de Muenster, Alemania. Es profesor titular del departamento de Filosofía de la Universidad Nacional, sede Bogotá.

Dirección electrónica: [arosasl@unal.edu.co](mailto:arosasl@unal.edu.co)

**María Andrea Arciniegas.** Psicóloga egresada de la Universidad Nacional de Colombia, sede Bogotá.

Dirección electrónica: [andreaarciniegas@gmail.com](mailto:andreaarciniegas@gmail.com)

**Esteban Caviedes.** Estudiante de Maestría en Filosofía de la Universidad Nacional de Colombia, sede Bogotá.

Dirección electrónica: [esteban.caviedes@gmail.com](mailto:esteban.caviedes@gmail.com)

**Alejandra Arciniegas.** Estudiante de Maestría en Filosofía de la Universidad Nacional de Colombia, sede Bogotá.

Dirección electrónica: [maarciniegas@unal.edu.co](mailto:maarciniegas@unal.edu.co)

**LA NEUROPSICOLOGÍA DEL JUICIO MORAL.**  
SOBRE LAS CAUSAS DE RESPUESTAS CONTRAINTUITIVAS  
A LOS DILEMAS MORALES\*

*Alejandro Rosas*  
*María Andrea Arciniegas*  
*Esteban Caviedes*  
*María Alejandra Arciniegas*  
Universidad Nacional de Colombia

**Introducción**

El debate sobre el papel de las emociones y de la razón en el juicio moral ha ocupado un lugar destacado en la filosofía moral. Hoy contamos con aportes de disciplinas científicas en ascenso que son bienvenidos para una filosofía que se nutre de las ciencias. Pensamos especialmente en un área de investigación creciente en neuropsicología cognitiva que es animada por preguntas de carácter filosófico, por ejemplo, la ya mencionada sobre el papel de las emociones en el juicio moral. ¿Son las emociones constitutivas del juicio moral o son más bien efectos que se producen una vez dictaminado el juicio, quizás como medios de los que se sirve el juicio para motivar al sujeto a una acción específica? En sus estudios pioneros, Greene *et al.* (2001, 2004) investigaron con imágenes de resonancia magnética funcional (IRMf) la activación neuronal de sujetos sanos al juzgar acciones descritas

---

\* Alejandra A. y Andrea A. agradecen el apoyo otorgado por la Facultad de Ciencias Humanas de la Universidad Nacional a través de la Convocatoria Orlando Fals Borda; A.R. y E.C. agradecen el apoyo de la Vicerrectoría de Investigación, código Hermes 16012; y A.R. agradece además el apoyo generoso de la John Simon Guggenheim Memorial Foundation.

en historias con dilemas morales. Encontraron que, enfrentados a una categoría de historias que llamaron ‘dilemas morales personales’ (DMP), las zonas emocionales del cerebro se activan más que cuando se trata de ‘dilemas morales impersonales’ (DMI). Más adelante intentaremos definir estas categorías. Por el momento, baste señalar que, en los DMP, el sujeto se enfrenta a un dilema hipotético en donde puede salvar varias vidas si le quita directamente la vida a un tercero. Las respuestas de los sujetos experimentales a los DMP coincidieron mayormente con normas deontológicas, que prohíben el sacrificio de inocentes sin importar las consecuencias; y entonces, los autores del estudio concluyeron que las emociones son importantes, quizás constitutivas, del juicio moral *deontológico*. Pero, a pesar de esto, no podían excluir la posibilidad de que las emociones fueran *efectos*, en lugar de *elementos constitutivos o causas*, de los juicios. Para resolver esta duda, otros investigadores utilizaron la misma metodología en pacientes con daño en la corteza pre-frontal ventromedial (Ciaramelli *et al.* 2007 y Koenigs *et al.* 2007). Los pacientes con daño ventromedial se caracterizan porque sus capacidades emocionales relevantes para la interacción social están seriamente comprometidas. La hipótesis de los estudios con pacientes decía que, si la emoción es constitutiva del juicio deontológico (y no un mero efecto del juicio), los pacientes, al no tener las emociones, no responderían de manera deontológica. Efectivamente, los pacientes tendieron a evitar la respuesta deontológica. Estrictamente hablando, la conclusión debió haber sido que esto mostraba que la emoción es constitutiva del juicio deontológico. Esta conclusión, por sí sola, hubiera sido interesante para un filósofo moral tentado por una perspectiva kantiana. Pues, según Kant, aunque los sentimientos morales están presentes cuando se emiten juicios acordes con el imperativo categórico, ellos no son constitutivos del juicio, sino tan sólo un efecto que la Razón Pura práctica ejerce sobre la sensibilidad del sujeto (Kant 2002 [1785], p. 17; AK 4:401). Kant sostuvo que el juicio moral tiene su origen en la Razón Pura y no en las emociones o inclinaciones del sujeto. Sostuvo esta tesis no sólo con miras a asentar la validez del juicio en la Razón. La Razón es para él también la causa y el origen de nuestra propensión a formular juicios morales y a interpretar las acciones e interacciones humanas como instanciando propiedades morales. Uno puede, con razón, sostener que Kant habría rechazado cualquier ingerencia de la neuropsicología en la cuestión de la validez de los juicios morales; pero si la ciencia prueba que la incapacidad para sentir ciertas emociones afecta el juicio moral de manera sistemática, este resultado tendría que afectar la posición de Kant sobre el carácter accesorio de las emociones en cuanto a las causas y el origen de nuestra propensión a ver el mundo en términos morales. Después de todo,

la neuropsicología sugiere precisamente que, si uno carece de las emociones apropiadas, no puede *ver* el mundo en términos morales. En cuanto a la relación entre explicación y justificación en la meta-ética, se trata sin duda de un asunto pertinente que no podemos tratar aquí (ver Rosas 2000)

Pero los investigadores formularon sus resultados de un modo que los involucró en un debate distinto: afirmaron que los pacientes mostraron una tendencia a juzgar dilemas morales desde un sistema normativo utilitarista. En ello seguían una sugerencia especulativa de Greene *et al.* (2004) sobre la existencia de dos módulos morales en el cerebro. Esta sugerencia conecta la investigación neuropsicológica con otro debate filosófico. En lugar del debate meta-ético (explicativo o descriptivo) sobre el rol causal de la emoción en la producción del juicio moral, ahora se trata de la rivalidad entre dos sistemas normativos: el utilitarismo y la deontología. ¿Pueden acaso las evidencias neuropsicológicas decirnos algo acerca de la rivalidad entre un sistema deontológico y uno utilitarista? No se trataría, de ningún modo, de darle a la neuropsicología la capacidad de dirimir los desacuerdos normativos entre aquellos sistemas o darnos luces sobre cuál es el moralmente correcto (Berker 2009). Pero la meta-ética tienen también un aspecto explicativo, y no se agota en tareas de tipo normativo y de justificación del juicio moral. La tesis de Greene asume, correctamente a mi modo de ver, que la neuropsicología podría incidir en el aspecto explicativo y aclarar quizás por qué las intuiciones morales varían en la población humana, y ubican, tanto a filósofos como a legos, a unos en el bando utilitarista y a otros en el bando deontológico. En principio, creemos que es posible que la neuropsicología dé una explicación de por qué unos sujetos se inclinan por respuestas utilitaristas y otros por respuestas deontológicas a los DMP. Escanear el cerebro de sujetos sanos mientras cumplen una tarea cognitivamente bien definida, o analizar con una perspectiva funcional las respuestas de los pacientes con daños localizados, puede aportar evidencias sobre la función que cumplen distintas subestructuras neuronales. Cruzando varios estudios similares y buscando una asignación coherente de funciones, podemos elaborar un mapa de correlaciones entre estructuras cerebrales y funciones cognitivas, y podemos inferir las funciones requeridas para una tarea observando los patrones de activación neuronal.

Pero si en principio la tesis de Greene cae dentro de la jurisdicción de la meta-ética explicativa, creemos que la hipótesis de los dos módulos cerebrales con funciones morales distintas – uno para la moral utilitarista y otro para la deontológica – es ficción: creativa, pero ficción al fin y al cabo. Para empezar, la hipótesis de los dos módulos reclama una plausibilidad evolutiva. El módulo deontológico sería ancestral, mientras que el utilitarista

sería reciente. Pero tanto el utilitarismo como la deontología son, además de propuestas de tipo normativo y de justificación de los juicios morales, intentos filosóficos por reconstruir los procesos cognitivos o incluso emocionales que subyacen a nuestras actitudes morales. Ambas propuestas son productos *culturales* altamente sofisticados y como tales, no son buenos candidatos para ser consideradas estructuras innatas, fijadas por la evolución. Es más probable que los procesos neuropsicológicos subyacentes al pensamiento moral tengan elementos de ambos sistemas y que los filósofos, con intenciones *normativas* más que *explicativas*, hayan intentado formular conjuntos de normas sistemáticamente puros y uniformes. En cualquier caso, las evidencias experimentales no hablan a favor de una hipótesis explicativa que postule dos módulos, sino que sugieren una explicación diferente, que apunta a una disfunción en la capacidad para la empatía (sección 2). Esta explicación concuerda mejor con teorías filosóficas plausibles –de corte Humeano– sobre el juicio moral (sección 3). En la sección 3 planteamos, además, una reflexión sobre el objetivo de estas investigaciones y proponemos que su objetivo es entender, neuropsicológica- y funcionalmente, lo que sucede en la mente de las personas cuando aprueban, *contra-intuitivamente*, acciones utilitaristas. Una vez que tenemos claro que éste es su objetivo, procedemos a evaluar la batería de dilemas de Greene desde el punto de vista de su idoneidad para evocar en los sujetos experimentales juicios que se puedan interpretar sin ambigüedad como juicios utilitaristas contraintuitivos. Para realizar esta evaluación, comenzamos proponiendo en la sección 4 las condiciones ideales que debe cumplir un dilema moral para presentar una opción *utilitarista* contra-intuitiva. Finalmente, planteamos en la sección 5 críticas a la batería de dilemas morales utilizada en estos estudios, precisamente porque muchos ítems de la batería no tienen en cuenta esas condiciones ideales. Paradójicamente, la batería es demasiado rica para alcanzar un objetivo relativamente restringido. Por tanto, no sería una mala idea ampliar el objetivo de estos estudios: en lugar de buscar tendencias al juicio *utilitarista* contra-intuitivo, enfocarse en el juicio moral *contra-intuitivo* sin más, e investigar sistemáticamente qué variables suscitan desviaciones en el juicio moral cuando los sujetos muestran disfunciones en sus capacidades empáticas.

### **¿Dos módulos para la moral en el cerebro?**

La hipótesis de los dos módulos fue sugerida por Greene *et al.* (2004) y luego defendida en Greene (2008). Propuso que la rivalidad entre moral utilitarista (p. ej., John Stuart Mill) y moral deontológica (por ej., Immanuel

Kant), que ha ocupado un lugar central en los debates de filosofía moral, puede tener una base neuronal. La moral utilitarista se apoyaría en subestructuras que evolucionaron más recientemente, mientras que la moral deontológica se asentaría en subestructuras emocionales ancestrales. Greene *et al.* (2004) no usaron el concepto de módulo. Lo utilizamos aquí de una manera amplia para referirnos a una subestructura cerebral dedicada a una función específica y con un origen selectivo propio, aunque la subestructura esté distribuida en distintas partes del cerebro. Greene (2007) defendió que su hipótesis de los dos módulos explica los resultados que otros investigadores obtuvieron con pacientes. Éstos habrían perdido el módulo deontológico y se guiarían sólo por el utilitarista. Serían utilitaristas en estado puro, producidos ya sea por enfermedad, como en la demencia fronto-temporal (Mendez *et al.* 2005; Lough *et al.* 2006) o por lesiones (aneurismas, accidentes en automóvil) como en los pacientes con daño ventromedial.

Los estudios con pacientes – y otros similares realizados con psicópatas (p. ej., Koenigs *et al.* 2012) o con sujetos sanos con puntajes altos en los cuestionarios que miden niveles de personalidad antisocial (Bartels y Pizarro 2011) – arrojan una inclinación al juicio utilitarista *contra-intuitivo* frente a una acción descrita en una viñeta que, para una mayoría, evoca una prohibición deontológica. Es conveniente presentar un par de viñetas que ilustran dónde se apoya la tesis de una inclinación anormal al utilitarismo. Las viñetas conocidas como *Transplant* y *Footbridge* pueden ilustrar el punto:

*Transplant:*

Un brillante cirujano de trasplantes tiene cinco pacientes, cada uno en la necesidad de un órgano diferente sin el cual cada uno de ellos moriría. Desafortunadamente, no hay órganos disponibles para realizar cualquiera de estas cinco operaciones de trasplante. Un viajero joven y saludable que está de paso por la ciudad viene a hacerse un chequeo de rutina. Mientras le hacen el chequeo, el doctor descubre que sus órganos son compatibles con todos sus cinco pacientes moribundos. Supongamos, además, que si el joven desapareciera nadie sospecharía del doctor. ¿Es moralmente admisible que el doctor use a este viajero para salvar las vidas de las cinco personas?

*Footbridge:*

Un tranvía fuera de control se dirige a toda velocidad hacia cinco personas que se encuentran en la vía, las cuales morirán si el tranvía sigue su curso actual. Usted está de pie junto a un hombre voluminoso en un puente peatonal que está sobre la vía. La única manera de salvar a las cinco personas es

empujar a este hombre del puente peatonal, dentro de la trayectoria del tranvía. ¿Es esto moralmente admisible?<sup>1</sup>

En ambos casos, una mayoría de los sujetos sanos rechaza la acción propuesta para salvar a 5 personas. Pero los pacientes, los psicópatas y los sujetos sanos con puntajes altos en los tests de personalidad antisocial se inclinan a aprobar esas acciones. La tesis de Greene es que tienen deficiencias en el módulo deontológico y contestan guiados por el módulo utilitarista.

La explicación bimodular, sin embargo, no es la única posible, si se tienen en cuenta las carencias que distinguen a estos grupos poblacionales. Habría dos hipótesis para explicar esa inclinación. La preferida por Greene *et al.* (2004) y Greene (2007, 2008) es que el juicio utilitarista es producido por un esfuerzo racional de deliberación crítica que se sobrepone a emociones morales de origen evolucionario antiguo, o que guía el juicio cuando esas emociones no se activan debido a daños neuronales. La otra hipótesis propone que el factor clave es simplemente una carencia de empatía, sin la intervención de ningún módulo utilitarista. Wiech *et al.* (2012) diseñaron un estudio para confrontar estas dos hipótesis. Utilizaron cuestionarios estandarizados para medir rasgos de personalidad en sus participantes sanos y correlacionarlos con sus respuestas utilitaristas en un experimento con dilemas morales personales (DMP). Los rasgos de personalidad medidos fueron la pobreza en empatía y la tendencia a la hiper-cognición, hipotéticamente relacionada con el deseo de buscar la opción más racional cuando se trata de tomar una decisión compleja. También escanearon el cerebro de los participantes mientras respondían a los DMP. Encontraron que ambos rasgos de personalidad se correlacionan positivamente con la tendencia a dar respuestas utilitaristas. Pero observaron una correlación entre la pobreza en empatía y ausencia de activación de la Corteza Cingulada Subgenual (CCS) que, según confirman otros estudios (Zahn *et al.* 2009), se activa cuando los sujetos sienten empatía o culpa. En cambio, no encontraron ninguna correlación entre valores altos para el rasgo psicológico de hiper-cognición y activación de la Corteza Prefrontal Dorso-lateral –el área que Greene *et al.* (2001) relacionaron con el juicio utilitarista en sanos– ni entre hiper-cognición y desactivación de la CCS. Según el razonamiento plausible de Wiech *et al.* (2012), sus resultados evidencian que el incremento del juicio utilitarista se debe a una pobreza de empatía en los sujetos que así la revelaron en el cuestionario de personalidad, mientras que los patrones de activación neuronal en los hiper-cognitivos no sugieren un esfuerzo de

<sup>1</sup> Las traducciones de estos dilemas y de los que siguen son nuestras. Los dilemas se traducen a partir del texto presentado en el suplemento a Greene *et al.* (2001). Ver nota 4.



deliberación racional y crítica. Así, propusieron que en éstos últimos la tendencia a responder de manera utilitarista se puede deber más bien a una dificultad para razonar sobre sus emociones. Esto pone en duda la teoría de los dos módulos; y además, sugiere que las emociones son esenciales para el juicio moral, y que la tendencia a decisiones utilitaristas *contra-intuitivas* en los DMP (sacrificar directamente a un tercero para salvar varias vidas) es una disfunción, que puede tener etiologías diferentes, en el procesamiento de las emociones. Este fue también el núcleo de la crítica planteada por Moll *et al.* (2007) a la hipótesis bimodular en Greene *et al.* (2004).

### **3. La filosofía moral Humeana y el objetivo de los estudios con dilemas morales**

Creemos que la teoría filosófica que sitúa las emociones de aprobación y desaprobación (sujetas a alguna forma de universalización) en el núcleo del juicio moral, y que históricamente tuvo su primera formulación en Hume, es la teoría correcta. Hay varias versiones de esa teoría, pero no es necesario hacer aquí una exposición de ninguna de ellas. Nos limitamos a recoger algunas observaciones de Jesse Prinz que son pertinentes en relación con los estudios con dilemas morales, sin por ello suscribir todos los aspectos de su teoría. Su teoría “emocionista” (Prinz 2007) plantea que las emociones son constitutivas de los juicios morales. Esta era, precisamente, la tesis que Ciaramelli *et al.* (2007) y Koenigs *et al.* (2007) se habían propuesto poner a prueba en sus estudios con pacientes ventromediales. En ella no hay cabida para dos módulos morales, pues todo juicio moral está constituido por emociones. Por ejemplo, la diferencia moral que vemos entre los crímenes contra niños y los crímenes contra adultos, o entre matar y dejar morir, se deriva de diferencias en la intensidad emocional que acompaña al juicio desaprobatario.

La teoría emocionista puede explicar por qué las personas tienden a decidir deontológicamente en unos dilemas (p.ej., *Footbridge*) y de manera utilitarista en otros (p.ej., *Trolley*)<sup>2</sup>, sin necesidad de postular módulos. En estos casos, es la intensidad de las emociones la que nos guía sobre qué juicios morales emitir. Los DMP como *Footbridge* generan emociones negativas fuertes, pues hay que quitarle la vida a una persona en una acción directa. La emoción aversiva generada por la idea de empujar a la persona a su muerte, así sea para salvar a más personas, es más poderosa que la emoción generada por la idea de salvar vidas. En términos de reglas, la regla

<sup>2</sup> La transcripción traducida de *Footbridge* se puede consultar arriba en la sección 2. *Trolley* es similar, pero el protagonista desvía el tranvía a una ruta alterna, donde sabe que otra persona será arrollada y morirá.

que prescribe “no mates” se apoya en una emoción fuerte, y la que ordena “es bueno salvar vidas” se apoya en una emoción más débil. La emoción fuerte toma el control y las personas deciden no empujar a la persona a una muerte segura, así dejen morir a otros por tomar esa decisión. En cambio, en *Trolley* la emoción positiva ligada a la idea de salvar vidas es más fuerte que la emoción aversiva hacia la muerte prevista al mover la palanca para desviar el tranvía (Prinz 2007 25); y en los casos en que la muerte se causa de manera indirecta, los números pesan (por eso se activan las áreas cerebrales de memoria de trabajo). No se trata de un proceso frío racional contra uno emocional, sino que se están enfrentando emociones fuertes contra emociones débiles. Prinz afirma que frente a una hipotética viñeta como *Trolley*, pero con la palanca muy cerca de la persona que moriría, los sujetos reaccionarán con una emoción similar a la que les produce *Footbridge*. O, al contrario, en una versión de *Footbridge* en donde se activa un mecanismo a distancia para lanzar a la persona del puente, la gente respondería de forma similar a como responden en *Trolley*. De hecho, esta última versión de *Footbridge* fue aplicada experimentalmente (Mikhail 2007) y confirmó en parte la predicción de Prinz, pues el porcentaje de aprobación subió, aunque no igualó los porcentajes de aprobación obtenidos con *Trolley*, lo cual es indicio de que hay otro elemento de la situación que dispara emociones fuertes de rechazo a la acción a distancia que lanza a la persona del puente. Todo depende, entonces, de las emociones que la viñeta y sus variables sean capaces de activar.

Estas intuiciones plausibles (algunas confirmadas), sugieren que frente a los dilemas morales las personas se guían por las emociones activadas por las viñetas. En realidad, el utilitarismo y la deontología, como sistemas normativos, son reconstrucciones filosóficas que pretenden capturar, en raciocinios formales distintos, lo que realmente sucede cuando producimos juicios morales. No es implausible que los procesos mentales que subyacen al juicio moral tengan aspectos de ambas reconstrucciones y que ninguna, en su forma pura, los capture completamente. Ni siquiera es claro que las dos reconstrucciones tengan que llegar, necesariamente, a juicios particulares distintos: es sólo que llegan allí desde distintos puntos de partida, y por vías diferentes.

Supongamos que la rivalidad entre ambos sistemas normativos no se exprese al nivel de los juicios particulares, sino sólo al nivel de la reconstrucción formal de los puntos de partida y de los raciocinios que nos llevan hasta ellos, lo cual es una conjetura plausible. Aún así, es un hecho que enfrentados al dilema *Footbridge*, por ejemplo, una minoría no despreciable emite un juicio *utilitarista contra-intuitivo*. Pero cuando describimos así

su juicio, no pretendemos, o no deberíamos pretender, que el utilitarismo como *sistema filosófico formal*, y sólo él, prescribe ese juicio. Se le llama “utilitarista” porque *superficialmente* parece justificado por un principio de bienestar agregado. Y se le llama *contra-intuitivo* porque el juicio se desvía de la norma: la mayoría se inclina por el juicio “deontológico”. Dado que “juicio utilitarista” es sólo una etiqueta superficial conveniente, mientras que “*contraintuitivo*” indica que los sujetos investigados se desvían de la norma, la investigación neuropsicológica podría describirse más propiamente como planteando esta pregunta: ¿qué sucede en la mente de las personas cuando se desvían de la norma y emiten un *juicio contra-intuitivo*? Obviamente, así planteada, la pregunta se dirige a una conducta judicial atípica sin asumir nada sobre la relación entre esa conducta y los sistemas formales que reconstruyen el pensamiento moral. La explicación a todas luces más plausible dice que, cuando los sujetos aprueban, *contra-intuitivamente*, una acción “utilitarista” (entrecorramos para indicar que la etiqueta se usa en sentido *superficial*) que se opone a una intuición moral ampliamente compartida como en *Footbridge*, hay algún tipo de disfunción subyacente, ya sea en la capacidad para la empatía, o en el procesamiento de los sentimientos producidos por ella. Desde este punto de vista, la disfunción afecta al juicio moral a secas, y parece profundamente equivocado atribuirle a quienes la padecen un punto de vista moral distinto, y específicamente utilitarista. En últimas, creemos que la explicación que postula dos módulos es víctima de dos confusiones: (i) equipara apresuradamente las reconstrucciones formales del pensamiento moral con puntos de vista normativos diferentes respecto de acciones particulares; (ii) confunde las etiquetas superficiales “deontología” y “utilitarismo”, aplicadas a juicios particulares, con la presunción de que esos juicios son propios y exclusivos de un sistema formal con el mismo nombre. Estas confusiones se pueden poner al descubierto del siguiente modo: supongamos que variamos el dilema *Footbridge* de modo que el número de personas salvadas al empujar al hombre voluminoso es del orden de miles. La empatía hacia las muchas personas que morirán podría ahora inclinar a personas sanas y con índices normales o altos de empatía a una respuesta “utilitarista”. En cambio, los pacientes o las personas pobres en empatía podrían responder igual (estadísticamente) que cuando mueren sólo cinco (como lo plantea el dilema original). Esto indicaría que ellos no razonan de modo utilitarista y que su problema no radica en emitir consistentemente juicios “utilitaristas”, sino en desviarse del juicio moral normal, que a veces es “deontológico” y a veces “utilitarista”. Y la raíz de su problema es la carencia en empatía (Moll y de Oliveira-Souza 2007). Obviamente, esto es una hipótesis, pero no es recomendable ignorarla.

Si consideramos posible que la pobreza en empatía lleve eventualmente a un juicio contra-intuitivo, no por ser “utilitarista”, sino por *no* serlo (en la comparación estadística con los controles), el foco de la investigación debe trasladarse a los juicios morales contra-intuitivos sin más. Ahora bien, si una línea de investigación se limitase a querer confirmar si un grupo de sujetos experimentales se inclina a juicios “utilitaristas” contra-intuitivos, eso no sería incorrecto, pero revelaría estrechez de miras. Los estudios aquí discutidos creen haber confirmado que los grupos poblacionales con daño neuronal o con rasgos de personalidad antisocial tienen tendencia al juicio “utilitarista” contra-intuitivo. Sin embargo, si bien hay evidencia clara de algún tipo de distorsión en su juicio moral, no hay buenas razones para describir la distorsión como una tendencia hacia juicios “utilitaristas”. La razón es muy sencilla: los autores de los estudios asumen *apresuradamente* que *todos* los dilemas de la batería aplicada presentan una opción de actuar que es calificable como “utilitarista”. La batería sí contiene historias de este tipo; y si todas fueran así, y los sujetos de esos grupos poblacionales mayoritariamente aprobasen la acción del protagonista, sería legítimo describir su respuesta como “utilitarista” y contra-intuitiva. Pero la gran mayoría de historias no es de ese tipo. Un examen de los ítems individuales puede comprobarlo. En la sección que sigue queremos dejar en claro que esta batería, tomada de Greene *et al.* (2001) y aplicada parcial o totalmente en casi todos los estudios es, paradójicamente, demasiado rica en variables para confirmar una hipótesis tan restringida.

### **Dilemas Morales Personales sensu stricto**

La batería de dilemas de Greene *et al.* (2001, 2004) contiene dilemas no-morales y dilemas morales personales (DMP) e impersonales. Aquí sólo nos vamos a ocupar de los DMP, que en esa batería suman 25 historias. Idealmente, los DMP invitan a los sujetos a decidir salvar o no la vida de varias personas a costa de una, cuando la víctima es identificable en el escenario y cuando el protagonista no la convierte en víctima simplemente como resultado de *desviar* una amenaza pre-existente. Esta segunda condición trata de capturar la diferencia entre la versión *personal* del dilema en donde el protagonista empuja a otra persona a su muerte para detener al tranvía (*Footbridge*) y la versión *impersonal* en donde el protagonista solo *desvía* el tranvía a una vía alterna donde, fortuitamente, se encuentra otra persona que previsiblemente morirá (*Trolley*). Hay otras formas de capturar esa diferencia: por ejemplo, podría decirse que en *Footbridge* se usa a una persona como medio y en *Trolley* la muerte de una persona es un efecto

colateral de salvar vidas. Más precisamente, podría decirse que la muerte en el primer caso tiene un rol causal en la salvación de cinco personas amenazadas, mientras que el segundo caso el rol causal es la desviación del tranvía, y no la muerte colateral. Para nuestro propósito no es importante decidir qué explicación captura mejor las diferencias entre DMP y DMI. Pues nos interesa señalar las condiciones *adicionales* que deben tener los DMP para ser *dilemas que ofrecen una opción utilitarista* y además *contra-intuitiva*. Estas condiciones adicionales no se cumplen en una gran mayoría de historias.

Dos viñetas que las cumplen a la perfección y pueden tomarse como paradigmas son *Transplant* y *Footbridge*. Ambas fueron introducidas originalmente por Foot (1967) y Thomson (1985) como experimentos mentales éticos. Los presentamos arriba en la sección 2. Estos dilemas cumplan las condiciones formuladas a continuación, y los llamamos por esa razón DMP *sensu stricto*:

- (1) la persona que toma la decisión de victimizar a alguien para salvar otras vidas no se beneficia de esa acción. De lo contrario, su acción podría tener una razón egoísta en lugar de una utilitarista, es decir, salvarse a sí misma, en lugar de maximizar el bienestar general.
- (2) la víctima no es responsable, o culpable, de la amenaza que se cierne sobre los otros. Pues, en caso contrario, no sería una víctima inocente y ya no sería claro que el dilema proponga una acción utilitarista contra-intuitiva.
- (3) la víctima inocente no debe estar en una situación sin salida, en la que morirá de todas formas, incluso si no es sacrificada; pues, de estarlo, sería difícil señalar la acción correcta y, aunque sería un dilema, no sería uno que propone una acción utilitarista y contra-intuitiva.

### **La riqueza de la batería de dilemas de Greene et al. (2001)**

**I)** La condición (1) dice que la persona que se enfrenta al dilema (el protagonista) no debe ser uno de los que se salva, ya que de serlo su razón para aprobar la intervención podría ser egoísta y no utilitarista. En la batería de veinticinco historias, sin embargo, encontramos seis historias en las que la acción del protagonista tiene sólo un beneficio personal y no hay ninguna otra persona que sea salvada. En un dilema moral debe haber, al menos en apariencia, dos normas morales en conflicto, que prescriben acciones opuestas. Pero en estas historias sólo se plantea un conflicto entre la norma moral y el egoísmo. Por ello, no merecerían el calificativo de dilemas morales. Los llamaremos *Dilemas Egoístas* y son: *Country Road*

(3), *Architect* (4), *Safari* (8), *Hired Rapist* (11), *Grandson* (12) y *Infanticide* (13).<sup>3</sup> Como ejemplo, miremos *Country Road* (3):

Usted está conduciendo por una carretera rural cuando escucha un pedido de auxilio que viene de unos arbustos junto a la carretera. Usted se detiene y encuentra a un hombre cuyas piernas están cubiertas de sangre. El hombre le explica que ha tenido un accidente mientras estaba en una excursión y le pide que lo lleve a un hospital cercano. Su inclinación inicial es ayudar a este hombre, quien probablemente perderá su pierna si no llega al hospital pronto. Sin embargo, si lo lleva al hospital, su sangre va a arruinar la tapicería de cuero de su auto. ¿Es apropiado que usted deje a este hombre al pie de la carretera con el fin de preservar su tapicería de cuero?

Obviamente la historia enfrenta al protagonista a una decisión moral, pero el conflicto es sólo entre la norma moral contra el egoísmo. *Country Road* no es un auténtico dilema moral y la aprobación de la acción del protagonista no podría describirse apropiadamente como una respuesta utilitarista.

Esto mismo sucede en los otros casos enunciados: en *Safari* el protagonista victimiza a un rehén que morirá de todas formas con el único fin de salvar su vida (fallan las condiciones 1 y 3); en *Hired Rapist*, el protagonista victimiza a su esposa con el fin de mejorar su relación con ella (falla la condición 1); en *Architect* el protagonista se deshace de su jefe para mejorar su vida profesional (falla la condición 1); en *Grandson* el protagonista gasta una “broma” letal a su abuela por haber sido privado por ella de la mesada (falla la condición 1); y, por último, en *Infanticide*, la joven se deshace de su bebé con el fin de evadir su responsabilidad como madre (falla la condición 1).

**II)** En otras diez historias, el agente se encuentra en una situación en la que salvando a otros también se salva o beneficia a sí mismo. No existe ninguna razón para creer que la persona actúa para salvar a los otros en lugar de a sí misma. Por esa razón es más apropiado considerarlos *Dilemas Mixtos*. Son las historias *Lifeboat* (5), *Hard Times* (6), *Smother for Dollars* (7), *Crying Baby* (9), *Plane Crash* (10), *Modified Lifeboat* (15), *Modified Safari* (17), *Submarine* (19), *Sacrifice* (22) y *Euthanasia* (25). Veamos el siguiente ejemplo (*Lifeboat*):

Usted está en un crucero cuando se inicia un incendio a bordo, y el barco debe ser abandonado. Los botes salvavidas están llevando más personas de

<sup>3</sup> Los números y nombres de las historias corresponden a la batería de Greene *et al.* (2001), que se puede consultar completa en <http://www.sciencemag.org/content/suppl/2001/09/13/293.5537.2105.DC1>

las que pueden cargar. El bote en el que usted se encuentra reposa tan bajo, que el agua está a punto de entrar y hundirlo. El mar empieza a agitarse y el bote comienza a llenarse de agua. A usted le parece que la única forma de evitar el hundimiento es comenzar a lanzar a otros pasajeros al agua comenzando con los ancianos, quienes son muy débiles para resistirse. ¿Es apropiado que usted lance algunos de sus compañeros pasajeros al agua para salvarse a usted mismo y a los pasajeros restantes?

Con su acción, el protagonista salvaría su propia vida además de la de otros pasajeros. Es difícil determinar cuál de las dos razones determina su acción y cuál determina la respuesta del participante que se ponga en sus zapatos. La simple posibilidad de que la razón egoísta determine la respuesta le quita la forma de un dilema moral. Con estas 10 y las anteriores 6 historias, tenemos 16 historias en las que la aprobación de la acción de un DMP podría ser evidencia de un juicio egoísta, en lugar de uno “utilitarista”.

III) Por otro lado, 5 historias incluyen otras variables potencialmente relevantes. En *Preventing the Spread* (14) y *Modified Preventing the Spread* (16), aunque el protagonista no se encuentra entre los que serían salvados, la víctima potencial amenaza explícita y creíblemente la salud de otras personas (tiene VIH positivo y quiere contagiar a otros), es decir, no es una víctima inocente (condición 2) pues tiene, a diferencia del sacrificado en *Footbridge*, responsabilidad directa en la amenaza que se cierne sobre las otras personas. Algo semejante se da en *Lawrence of Arabia* (20), en donde se trata de ejecutar a un asesino, o a un chivo expiatorio de su misma tribu, para evitar una guerra; y en *Modified Bomb* (18), se tortura al hijo de un terrorista en su presencia para disuadirlo de detonar una bomba en un lugar concurrido. En *Sophie's Choice* (21), en cambio, el hijo sacrificado moriría incluso sin la intervención de la protagonista para salvar a su otro hijo, es decir, no se cumple la condición (3). En este caso parece obvio que no hay ninguna acción deontológicamente correcta, y por tanto no se trata de una opción utilitarista contra otra deontológica. En suma, estos cinco dilemas presentan variables que impiden describir el sacrificio de la víctima como “utilitarista” y contra-intuitivo.

En total, 21 de las 25 historias que según Greene *et al.* son DMP, contienen variables que no están relacionadas con un juicio utilitarista contra-intuitivo. Sólo 4 historias corresponden a un DMP *sensu stricto*: *Footbridge* (1), *Transplant* (2), *Vitamins* (23) y *Vaccine Test* (24).

En conclusión, la batería contiene historias con variables ajenas al conflicto entre deontología y utilitarismo y no permite poner a prueba una hipótesis limitada a ese conflicto, como proponen los estudios con dilemas morales aquí discutidos, a saber, que algunos grupos poblacionales tienen

una inclinación al juicio “utilitarista” contra-intuitivo. Por ejemplo, en las seis historias que llamamos *Dilemas Egoístas*, y que los sujetos sanos de Greene desaprueban sin vacilación (ver el re-análisis de los datos de Greene *et al.* 2001, en McGuire *et al.* 2009, quienes llamaron a estas historias “no-dilemáticas”), no hay diferencia con los pacientes ventromediales estudiados por Koenigs *et al.* (2007). El 100% de estos pacientes rechazaron la acción, es decir, se comportaron igual que los sujetos sanos de Greene *et al.* (2001).<sup>4</sup> Por esa razón Koenigs *et al.* (2007) las clasificaron como de “bajo conflicto”, aunque hubiese sido más apropiado decir que no son dilemas morales. Los pacientes parecen tener claro que las acciones egoístas propuestas en esos dilemas deben ser rechazadas. Pero llama la atención que los pacientes de Koenigs *et al.* (2007) rechazaron en un 100% la acción en *Transplant* y en *Plane Crash*. Esta última contiene un elemento de canibalismo practicado sobre un niño, que podría dar la explicación se su rechazo. Este es un ejemplo de cómo variables ajenas al conflicto entre respuesta deontológica y respuesta utilitarista pueden determinar las respuestas. El caso de la respuesta unánime desaprobatoria de los pacientes a *Transplant* es, en cambio, un contraejemplo a la hipótesis del paciente utilitarista. *Transplant* es un caso claro de DMP *sensu stricto*, en donde la opción “utilitarista” es contra-intuitiva. Los pacientes optaron por la acción deontológica sin excepción.

Esto debería plantear dudas sobre la literatura que los describe como teniendo una inclinación al utilitarismo sin más, y sugerir que se debe investigar la incidencia de las otras variables en sus respuestas. Sería interesante estudiar si sus respuestas a los dilemas mixtos, por ejemplo, o a los dilemas con culpables o con miles de personas por salvar, difieren significativa y sistemáticamente de las respuestas de los controles sanos. Esto liberaría a la investigación neuropsicológica de insistir en un enfoque restringido y casi obsesivo con el juicio “utilitarista” contra-intuitivo y podría arrojar resultados más iluminadores sobre las deficiencias en el juicio moral en poblaciones con disfunciones en su capacidad para la empatía.

---

<sup>4</sup> Agradecemos a los autores del estudio habernos facilitado los datos originales.



## Referencias Bibliográficas

- Bartels, D. M., & Pizarro, D. A. (2011): "The mismeasure of morals. Antisocial personality traits predict utilitarian responses to moral dilemmas", *Cognition* 121(1): 154-161.
- Berker, Selim (2009): "The Normative Insignificance of Neuroscience," *Philosophy and Public Affairs* 37: 293-329.
- Ciaramelli, E., Muccioli, M., Ladavas, E. y di Pellegrino, G. (2007): "Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex", *Social Cognitive and Affective Neuroscience* 2(2): 84-92.
- Foot, Philippa (1967): "The Problem of Abortion and the Doctrine of Double Effect," *Oxford Review* 5: 5-15
- 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.
- Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M. y Cohen, J.D. (2004): "The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44: 389-400.
- Greene J.D. (2007): "Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains", *Trends in Cognitive Science* 11(8): 322-323.
- Greene, J.D. (2008): "The secret joke of Kant's soul", en W. Sinnott-Armstrong (ed.), *Moral psychology: The neuroscience of morality*, Cambridge: MIT Press, pp. 35-79.
- Kant, I. (2002 [1785]). *Groundwork for the metaphysics of Morals*. trad. por Allen Wood. New Haven: Yale University Press.
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M. y Damasio, A. (2007): "Damage to the prefrontal cortex increases utilitarian moral judgments", *Nature* 446 (7138): 908-911.
- Koenigs, M., Kruepke, M., Zeier, J. y Newman, J. P. (2012): "Utilitarian moral judgment in psychopathy", *Social Cognitive and Affective Neuroscience* 7(6): 708-714.
- Lough, S., Kipps, C.M., Treise, C., Watson, P., Blair, J.R., Hodges, J.R. (2006): "Social reasoning, emotion and empathy in frontotemporal dementia", *Neuropsychologia* 44, 950-958.
- McGuire, J., Langdon, R., Coltheart, M., y Mackenzie, C. (2009): "A reanalysis of the personal/impersonal distinction in moral psychology research", *Journal of Experimental Social Psychology* 45(3): 577-580
- Mendez, M.F., Anderson, E. y Shapira, J.S. (2005): "An investigation of moral judgment in frontotemporal dementia", *Cognitive and Behavioral Neurology* 18, 193-197
- Mikhail, J. (2007): "Universal moral grammar: theory, evidence and the future", *Trends in Cognitive Sciences* 11(4): 143-152
- Moll, J., & de Oliveira-Souza, R. (2007): "Moral judgments, emotions and the utilitarian brain", *Trends in Cognitive Sciences* 11(8): 319-321

- Prinz, J. (2007). *The Emotional Construction of Morals*. New York: Oxford University Press.
- Rosas, A. (2000). “Explicación y justificación. Hacia el naturalismo en la filosofía moral.” *En Mentes Reales. La Ciencia Cognitiva y la Naturalización de la Mente*. Bogotá: Universidad Nacional de Colombia/ Siglo del Hombre Editores, pp. 101-118.
- Thomson, J. J. (1985): “The Trolley Problem”, *The Yale Law Journal* 94/6: 1395-1415.
- Wiech K., Kahane G., Shackel N., Farias M., Savulescu J., Tracey I. (2013): “Cold or calculating? Reduced activity in the subgenual cingulate cortex reflects decreased emotional aversion to harming in counterintuitive utilitarian judgment”, *Cognition* 126(3): 364-72.
- Zahn R, de Oliveira-Souza R, Bramati I, Garrido G, Moll J (2009): “Subgenual cingulate activity reflects individual differences in empathic concern”, *Neurosci Lett* 457:107–110.