

DOI: 10.25100/pfilosofica.v0i50.8725

## EL VALOR DE LOS DILEMAS MORALES PARA LA TEORÍA DE LAS DECISIONES

**Fabio Morandín-Ahuerma**

Complejo Regional Nororiental,

Benemérita Universidad Autónoma de Puebla, Puebla, México.

### **Resumen**

*En este artículo se analiza la teoría para la toma de decisiones y se contrasta la valía de los dilemas morales para explicar los mecanismos deliberativos. Se sostiene que una estrategia de decisión exitosa es aquella que es capaz de realizar los movimientos racionales, adaptativos y necesarios para llegar a un fin programado. Una aspiración de las éticas normativas y descriptivas es elaborar una teoría de la decisión práctica, sin importar los análisis racionales del modo en que se llegue al fin. El artículo plantea los procesos relacionados al cambio y a la retroalimentación sin encasillar la teoría de las decisiones en posturas deontológicas, utilitaristas o emocionales. Se examina la valía de los dilemas morales y algunas críticas recurrentes y se concluye que, personales o impersonales, los dilemas parecen no explicar la complejidad de los procesos deliberativos.*

**Palabras clave:** *teoría de las decisiones; dilemas morales; cerebro moral; retroalimentación; dilema del tranvía.*

**Cómo citar este artículo:** Morandín-Ahuerma, F. (2020). El valor de los dilemas morales para la teoría de las decisiones. *Praxis Filosófica*, (50), 187-206. doi: 10.25100/pfilosofica.v0i50.8725

**Recibido: 1 de mayo de 2019. Aprobado: 15 de junio de 2019.**

*Praxis Filosófica*, No. 50 enero - junio 2020: 187 - 206

ISSN (I): 0120-4688 / ISSN (D): 2389-9387

# The Value of the Moral Dilemmas for the Theory of Decision-Making

*Fabio Morandín-Ahuerma*<sup>1</sup>

## *Abstract*

*This article analyzes the theory for decision making and contrasts the value of moral dilemmas to explain the deliberative mechanisms. It is argued that a successful decision strategy is one that is capable of performing the rational, adaptive and necessary movements to reach a programmed end. It suggests that the processes related to change and feedback can be useful but insufficient explanations if the theory of decisions is framed to deontological, utilitarian or emotional postures. The value of moral dilemmas and some recurrent criticisms are examined and it is concluded that, personal or impersonal, the dilemmas do not seem to explain the complexity of the deliberative processes.*

**Keywords:** *Decision-Making Theory; Moral Dilemmas; Moral Brain; Feedback; Trolley Dilemma.*

---

<sup>1</sup> Es Doctor en Filosofía por el Instituto de Filosofía de la Universidad Veracruzana (México). Realizó una estancia postdoctoral en el Centro de Investigaciones Filosóficas de Buenos Aires en el programa de Neuroética. Actualmente es Profesor Investigador de la Benemérita Universidad Autónoma de Puebla (México). Premio *Arte Ciencia Luz* 2015. Entre sus artículos se encuentran: (2019) ¿Quién mató a Elaine? Autos robos y toma de decisiones, en *Elementos*. (2019) Deontología y derechos humanos: ¿Por qué no se debería desviar el tranvía?, en *Doxa*. (2018) Una reivindicación de mente y voluntad frente al conductismo lógico de Ryle, en *Educación y Desarrollo*, entre otras publicaciones. Actualmente se encuentra trabajando en el proyecto de investigación: «Estudio comparativo sobre teorías de correlatos biológicos en la toma de decisiones morales» (BUAP-VIEP, ID: 00687).

**ORCID:** 0000-0001-6082-2207 **E-mail:** fabio.morandin@correo.buap.mx

# EL VALOR DE LOS DILEMAS MORALES PARA LA TEORÍA DE LAS DECISIONES\*

*Fabio Morandín-Ahuerma*

Complejo Regional Nororiental, Benemérita Universidad Autónoma de  
Puebla, Puebla, México.

## **I. Introducción**

### **I. 1. Definiciones**

La definición de teoría de la toma de decisiones comúnmente aceptada [que en adelante nos referiremos a ella por sus siglas TDM, del inglés, *Theory of Decision Making*] se refiere a los procesos adaptativos de acuerdo a las variaciones que presenta cada contexto para realizar el acto de decidir. Para una adecuada o ventajosa TDM el sujeto está consciente de que tiene múltiples opciones y, de las variables, escoge la opción que mejor se adapta a sus propios objetivos, tanto de corto como de mediano y largo plazo. Para ello, se debe inhibir tomar una decisión inmediata o reactiva, ya que nos puede conducir a resultados inesperados, aleatorios o inciertos (Coutlee y Huettel, 2012; Kaplan y Mikes, 2012; Bland, 2012). Sin embargo, algunos escenarios nos obligan a tomar decisiones de esa naturaleza.

La TDM *exitosa* es aquella capaz de distinguir los mejores valores entre múltiples opciones, sacar consecuencias por adelantado, descartar supuestos erróneos o prejuicios sin fundamento y, a partir de ello, definir la estrategia de acción más adecuada a los objetivos preestablecidos o a las

---

\*Agradezco las observaciones realizadas por los revisores externos en el proceso anónimo par ciego llevadas a cabo en este trabajo, las cuales sirvieron para eliminar erratas y enriquecer al texto.

necesidades por cubrir. Ante la incertidumbre, se requiere de una selección flexible de los distintos cursos de acción y comportamiento (Coutlee y Huettel, 2012; Buzzell y Slater, 1962). La TDM consiste en un proceso recursivo de preparación, análisis inmediato o mediato, consciente o no; el paso necesario a la acción o inacción y; por último, la retroalimentación y evaluación de resultados para dar inicio nuevamente al ciclo (MacKinnon, Fairchild y Fritz, 2007; Coutlee y Huettel, 2012).

La creencia de que el camino aparentemente más corto, la idea más simple o la teoría con el menor número de elementos representa ventajas comparativas frente a otras opciones más complejas, no siempre resulta eficiente (Morin y Pakman, 1994) por el contrario; evitar las explicaciones simplistas, las aseveraciones gratuitas y las conclusiones precipitadas, puede ser el comienzo de un proceso más elaborado, rico y adaptativo, según el contexto, a las variables que se presenten y a los objetivos que se persigan (Morin y Pakman, 1994).

Tres elementos complejos en la TDM: primero, proyectar distintos cursos de acción que deben valorarse de acuerdo a su competitividad, a veces, en conflicto unos con otros (Doz, 1996); en segundo lugar, evaluar los recursos de presupuestación para atender las demandas, especialmente en aquellas tareas que pueden ser recurrentes y que puedan servir de referencia para la creación de escenarios prospectivos y; en tercer lugar, la observancia de reglas mínimas de comportamiento ético, necesarias para que nuevos ciclos de interacción puedan realizarse (Doz, 1996; Coutlee y Huettel, 2012; Katz, 1964).

A veces es necesario cambiar viejos patrones de conducta o visiones preestablecidas porque imposibilitan el *input* de nuevos estímulos o de innovadoras formas de apreciación. Incluso a nivel perceptual, como parte de la TDM, se debe cambiar la *forma de ver* un esquema o problema (Kornmeier y Bach, 2012) para resolverlo. Un ejemplo literal de lo anterior es la capacidad o incapacidad para ver figuras tridimensionales *escondidas* en las imágenes estereogramas de puntos aleatorios (RDS) (Hinton y Becker, 1990). Se requiere de cierto entrenamiento para poder *acceder* a la imagen que está *dentro*, y hay muchas personas que jamás logran visualizarlas. Esto debido a que, por más que se les explique con palabras cómo deben ir *rompiendo* el punto focal inmediato, esto es, haciendo ligeramente con los ojos un *bizco*, no les es posible cambiar su esquema de visión bidimensional. Del mismo modo, se requiere cierto tipo de *inteligencia* para poder aplicar una apreciación distinta a la que, por comodidad, se ha tenido por valiosa a lo largo de la vida y emprender rutas alternas (Kornmeier y Bach, 2012). Hacemos cosas una y otra vez de la misma manera, aun cuando no estamos

recibiendo los *outputs* o resultados esperados. Una estrategia para la TDM exitosa es aquella que es capaz de realizar movimientos adaptativos constantes, pero controlados racionalmente sin perder de vista el objetivo.

## **I. 2. La retroalimentación**

La retroalimentación es el proceso en la TDM que surge como resultado del aprendizaje basado en problemas y su colección de experiencias (Warren, Dondlinger, McLeod, y Bigenho, 2012). Hay retroalimentación cuando el agente es consciente de que a cada curso de acción corresponde un resultado diferenciado. En la medida en que se pueden prever esos resultados, sobre todo si son los deseados, se genera un bagaje empírico que permite sacar un aprendizaje que, a su vez, sirve para futuros ciclos de acción (Coutlee y Huettel, 2012; Bhakthavatsalam y Cartwright, 2017).

La retroalimentación nos diferencia del resto de las especies. Se puede aplicar en casi todos los animales un conductismo basado en premios y castigos, estímulos y respuestas, pero el ser humano es el único capaz de analizar, sintetizar y utilizar provechosamente la retroalimentación para la construcción de estrategias diferenciadas (Bindra, 1978). La conciencia de que se tiene cierto control sobre los actos y sobre los protocolos que el cerebro establece en el proceso de TDM es privativo del hombre.

Si bien se puede ser reactivo, esa respuesta puede ser modificada por la retroalimentación pasada, construcción del presente y visualización del futuro; de ese modo, el ser humano es capaz de adecuar sus expectativas de forma deliberada y actuar en consecuencia (Coutlee y Huettel, 2012; Gintis, 2007; Stajkovic y Luthans, 1998).

## **I. 3. ¿Quién decide?**

La TDM no ha logrado integrar un marco explicativo común, y aun cuando se han tenido hipótesis generales más o menos compartidas, aun no se logra explicar convincentemente o construir un marco explicativo común, sobre cómo funcionan algunas áreas del cerebro comprometidas en la TDM y, especialmente, en la indecisión.

Cada cerebro es diferente y las generalizaciones de patrones comunes de actividad neural pueden ser propositivas, incluso demostrativas de algunas teorías, más no por ello conclusivas (Coutlee y Huettel, 2012, Verplaetse, 2014).

Se cree que la corteza prefrontal (PFC) cuenta con un sistema de modulación ante otros sistemas afectivos, inmediatos, emocionales y de recompensa (Drevets, Price y Furey, 2008; De la Fuente y Leefmans, 2015; Bechara, Damasio, Damasio y Lee, 1999). Un reto significativo para la TDM

y para los neurocientíficos que observan la actividad neuronal a través de imágenes de resonancia magnética (IRM), lo constituye una explicación convincente de los procesos de autocontrol frente a la recompensa inmediata (Coutlee y Huettel, 2012; Van den Bos y Güroğlu, 2009; Papageorgiou, Sallet, Wittmann, Chau, Schüffelgen, Buckley y Rushworth, 2017).

Lo anterior se refiere, por ejemplo, al manejo de las adiciones: a la inhibición de los mecanismos automáticos de recompensa, como en el caso de quien se pone a dieta (Berthoud, Münzberg, y Morrison, 2017) o al mantenimiento de la motivación (Maslow, 1943) o la voluntad para seguir haciendo una rutina de ejercicio por largos periodos (Coutlee y Huettel, 2012). Ni hablar de las toxicomanías. Los mecanismos de auto-mantenimiento del equilibrio homeostático del cuerpo no siempre envían las señales somáticas adecuadas o consistentes con los proyectos de largo alcance de *su propietario* (Higgs, 2016) por el contrario, el cuerpo, como si fuera *un extraño*, quiere satisfacer de inmediato *sus* necesidades básicas sin importar que esa *satisfacción* sea el origen de su propio deterioro y, a la larga, de su auto-destrucción. No estamos sugiriendo una dualidad mente-cuerpo o un homúnculo racional en el cerebro y un potro desbocado en el cuerpo, como lo imaginaba Platón (Platón, 2013) y tantos otros dualistas hasta nuestros días.

Al parecer, está aún lejana la comprensión total de los mecanismos cerebrales que se extienden por el cuerpo a través del sistema nervioso central y la explicación convincente de las funciones intero y exteroceptivas en la TDM (Singer y Tusche, 2014; Northoff y Bermpohl, 2004). Aun con los adelantos técnicos y tecnológicos con que cuenta las ciencias de la conducta, no se tiene un diagnóstico preciso de los intrincados nudos y funciones involucradas en las respuestas conductuales somáticas aprendidas o pre-establecidas, frente a nuevos estímulos externos (Armél y Ramachandran, 2003; Damasio, 1996).

No queda esclarecida cuál sería la función de la PFC para echar a andar los mecanismos inhibitorios frente a ciertas respuestas automáticas (Ridderinkhof, Van Den Wildenberg, Segalowitz y Carter, 2004; Coutlee y Huettel, 2012). Algunos estímulos parecen, o son, *irresistibles* para la persona, e inexplicables para una teoría racional de las decisiones (Bernhard, Chaponis, Siburian, Gallagher, Ransohoff, Wikler y Greene, 2016). La pregunta: *¿Quién manda aquí?* (Gazzaniga, 2012) es atinente frente a las disyuntivas que presentan los problemas reales de agencia (Pereboom, 2014): pensamientos obsesivos, acciones compulsivas y manías (Hall, Dhillá, Charalambous, Gogos y Karayiorgou, 2003) que las posiciones teóricas dualistas, mente y cuerpo o cuerpo y cerebro, extravían el foco del problema.

De hallarse una explicación satisfactoria a lo *irresistible* y su control, el camino de la cura se abriría, con resultados positivos hacia una amplia gama de aplicaciones para el fortalecimiento del autocontrol (Wisner, Jones y Gwin, 2010) y con ello hacia la solución final a miles de padecimientos y sufrimientos, debido a una *voluntad débil* y a las dependencias física y mental que hacen del ser humano esclavo de sus propios *deseos* en contradicción. Quiere dejar de hacer algo, pero no puede dejar de hacerlo. Tabaquismo, obesidad, diabetes no congénita, alcoholismo, dependencia a los fármacos neurológicos legales e ilegales, opioides, anfetamínicos y un largo etcétera, podrían ser quitados de la dieta diaria, salvando, literalmente, millones de vidas (Berthoud, Münzberg, y Morrison, 2017).

Por lo anterior, una aspiración de las éticas normativas y descriptivas es elaborar una TDM que *sirva* para tomar siempre la *mejor decisión*, sin importar los análisis científicistas del *modo* en que se llegue a ella.

## II. Racionalidad y emociones en el cerebro

### II. 1. Racionalidad

La TDM desde ha tratado de ceñirse a un solo esquema: el enfrentamiento entre el deontologismo, el consecuencialismo y el emotivismo pero no estamos seguros si deba ser el objetivo útil del estudio filosófico de la ética, en cuanto procesos cognitivos de decisión.

Si se decide de manera pragmática, se debe optar por el mayor beneficio (Mill, 1980); si se decide por principios, se debe cumplir con el deber (Kant, 2018); si se decide intuitivamente, la emoción imperante decidirá el curso de acción (Hume, 2005); si se decide por estímulo-respuesta, el cerebro podría estar decidiendo sin que sea racional o consciente del porqué se hace lo que se hace (Libet, 1985).

No hay recetas para la razón práctica, actuar de manera fundamentalista o deontológica, o actuar de manera consecuencialista o pragmática por principio, seguramente traerá efectos colaterales no deseados para quien lo practique ciegamente. Tampoco sería aceptable sostener que: *matar de vez en cuando podría ser necesario si la causa lo amerita*; pero tampoco el cumplimiento de *todos* los cánones puede ser sostenido al absurdo, como en el ejemplo del asesino que pregunta por su futura víctima y al cual, según Kant, no le deberíamos mentir (Kant, 1993). Pero tampoco se puede actuar permanentemente sin otro principio más que el de la economía y la potenciación de la ganancia, sin ningún marco normativo, aunque este sea el criterio que hoy prevalezca.

## II. 2. La intuición

Por siglos, y aun recientemente, se ha considerado que el proceso de toma de decisiones y construcción del juicio moral está basado en procesos racionales y cognitivos, en donde la deliberación de un precedente permite, por necesidad lógica, alcanzar una conclusión (Kohlberg, 1969). Este consecuente se infiere de las premisas, a partir de la deliberación, este es el principio de la toma de decisión racional.

Sin embargo, esta visión racionalista y cognitiva ha sido puesta en duda y cada día es mayor el número de investigadores que considera que la intuición y la emoción podrían jugar un papel más importante en la decisión y, por ejemplo, construcción del juicio moral, de lo que se pensaba (Haidt, 2001; Greene, Nystrom, Engell, Darley y Cohen, 2004). También afirman que el proceso de cognición y racionalización pertenecen a un *segundo momento* en el que se construye una justificación, *a posteriori* de la decisión intuitiva-emotiva ya tomada (Haidt, 2001). Haidt sostiene que, primero se decide intuitiva y emocionalmente, y después se justifica dicha decisión.

Hasta Kohlberg (1969) los juicios morales partían de modelos racionalistas de decisión abstractos; después emergieron con fuerza otras perspectivas, como el intuicionismo social (Haidt, 2001) basado en respuestas *rápidas* desde la emoción, por ejemplo, en un dilema moral.

Greene, Nystrom, Engell, Darley y Cohen (2004) descubrieron que el proceso racional y el proceso emocional incluso compiten en algunas decisiones. Trabajan juntas a veces, pero a veces entran en conflicto, uno y otro *sistema* de decisión (Evans y Stanovich, 2013). El viejo adagio de pensar con la cabeza y no con el corazón, o con el hígado.

## III. Los dilemas morales

Un instrumento metodológico comúnmente utilizado para la investigación del proceso de toma de decisión y para la construcción de la TMD ha sido distintos dilemas morales. La mayoría son artificialmente contruidos y se cree que, cuando el participante responde en un ambiente controlado, también refleja lo que sucede o sucedería en la *vida real*. La segunda parte de este artículo trata de enunciar algunos de estos dilemas y analizar si el valor de sus resultados puede o no ser generalizable para una teoría, especialmente del juicio moral.

### III. 1. Evatlo

Existen muchos tipos de dilemas, los primeros datan de la Grecia antigua, como el Dilema de Evatlo (Aulus Gellius, 1927) quien era un joven aristócrata que quería ser abogado y estaba bajo el patrocinio de Protágoras

de Abdera, el gran sofista y experto en retórica. Ambos habían acordado que el alumno pagaría al maestro la cantidad de 10 mil dracmas por su tutoría.

Sin embargo, también estipularon que no pagaría Evatlo a su maestro Protágoras hasta haber ganado su primer juicio. Sin embargo, dado que nadie contrataba los servicios de Evatlo, Protágoras cansado de esperarlo, decidió interponer una demanda sabiendo que aquel enfrentaría el siguiente dilema:

—Si Evatlo perdía el juicio, los jueces obligarían a pagarle a Protágoras; y si ganaba Evatlo, también tenía que pagarle, pues se habría cumplido la condición de ganar el primer caso y, por tanto, tendría que pagarle a su mentor.

Lo que Protágoras no previó fue que Evatlo presentaría un contradilema:

—Si los jueces consideraban que el alumno tenía la razón, entonces no tenía que pagarle al maestro y, si perdía, tampoco tenía que pagarle, pues no habría ganado su primer juicio y, por lo tanto, no se habría cumplido la cláusula.

Y fue así como el aventajado alumno evadió pagarle al maestro. La pregunta es: ¿Quién tenía la razón? ¿Protágoras o Evatlo?

### III. 2. Trolley

A finales de los años 60s del Siglo XX (Foot, 1967) y a partir de mediados de los 80s (Thomson, 1985) se conformaron algunos dilemas con los que se han realizado pruebas controladas para analizar los procesos racionales e irracionales de la toma de decisiones.

Los dilemas morales han sido divididos básicamente en dilemas morales personas (DMP) y dilemas morales impersonales (DMI) (Rosas, Arciniegas, Cavides y Arciniegas, 2014) en los primeros, el sujeto de estudio debe verse así mismo actuando y tomando decisiones arriesgadas; en el segundo caso, la persona es un espectador y solo se le cuestiona sobre lo que *otros deberían* hacer en ese caso (Rosas, Arciniegas, Cavides y Arciniegas, 2014). En los dilemas personales, por ejemplo, se infringe un daño corporal de forma directa a un individuo o a un grupo de personas de manera intencional, y ese daño tampoco es colateral como resultado de la búsqueda de un mal menor (Foot, 1967). En el DMP se realiza una acción intencional, que puede contener un contacto físico, causar lesiones o la muerte y es la persona misma quien realiza la acción i.e. *agencia*. En el DMI es otra persona quien realiza la acción y por lo regular, utiliza su racionalidad instrumental o técnica, así que el sentido hipotético de la situación permite al sujeto de estudio tomar cierta distancia. Esto parece un hecho menor, pero en los estudios realizados representa un significativo cambio de enfoque del participante y, por lo tanto, de los resultados (Rosas, Alfonso, Arciniegas y Gómez, 2013).

El clásico dilema del tranvía tiene dos versiones: una personal y otra impersonal. La primera debe verse *uno mismo* (personal) manejando un tren sin frenos que se encuentra, más adelante a cinco trabajadores que no se percatan del convoy que se acerca y que serán atropellados a menos que gires el mando hacia otra vía en donde hay un solo trabajador: —¿*Matarías* a uno para salvar a cinco? Esta es la presentación del dilema de manera personal (Thomson, 1985).

En cambio, una versión impersonal es que hay un tranvía sin frenos, el conductor se ha desmayado y hay una persona parada en la aguja de cambio de rieles, puede presionar un botón para que cambie la ruta, igualmente, de un lado hay cinco personas y del otro lado una persona: —¿Debería el testigo matar a uno para salvar a cinco? (Thomson, 1985).

Como puede verse el planteamiento es distinto, en un caso es el propio actor quien *mata* a uno o a cinco, en cambio en el segundo, es *otro* quien realiza la acción y el sujeto de estudio es tan solo un espectador.

Para que hacer más clara esta distinción, otro dilema, ya clásico, servirá de ejemplo:

196

### III. 3. Fat

En lugar de accionar un botón o palanca que haga el cambio de vías, ahora *tú* debes empujar a un hombre lo suficientemente grande sobre la vía desde un puente peatonal para, de esa manera, frenar el tranvía y evitar que atropelle a los cinco trabajadores. El dilema sigue siendo el mismo: —¿Es lícito sacrificar una vida para salvar a cinco? (Thomson, 1985).

El planteamiento y las emociones que causan cada dilema son diferentes y, por tanto, los resultados también difieren. En realidad, este segundo dilema no es muy difícil porque la mayoría de las personas simplemente dicen que *no lo harían* (Greene, Nystrom, Engell, Darley y Cohen, 2004; Kamm, 2015) no empujarían a alguien para salvar la vida de otros porque es un homicidio intencional, y a casi nadie le gusta verse así mismo matando a un inocente, mucho menos con sus propias manos, aun cuando se tenga el mejor motivo e intención para hacerlo.

En el caso del primer dilema, en cambio, varían las respuestas dependiendo de la perspectiva del planteamiento que se haga: muchas personas están de acuerdo en que *debería aquella persona* accionar el botón para salvar a cinco (Greene, Nystrom, Engell, Darley y Cohen, 2004; Kamm, 2015; Thomson, 1985). Esto es, resulta *lícito* accionar una palanca u oprimir un botón; en cambio parece detestable para una mayoría empujar a una persona a su muerte. Sin embargo, hay una cierta inconsistencia en ello, pues el resultado es el mismo (Lanteri, Chelini y Rizzello, 2008).

### III. 4. Sofia

Otro dilema, este considerado *difícil*, es el llamado Dilema de Sofia: Un soldado nazi en un campo de concentración le dice a Sofia, una madre polaca, que matará a sus dos hijos a menos que ella escoja a uno, salvándole la vida al otro: —¿Debe Sofia escoger a uno de ellos o dejar que el soldado se lleve a los dos? (Styron, 1979).

Algunos dilemas no miden los resultados de lo que *se debe* escoger, porque la mayoría de las personas se sienten tan incómodas y frustradas al escucharlos que los investigadores se han concentrado más en los tiempos amplios de reacción que en las respuestas mismas (Evans, Dillon y Rand, 2014).

Esto es, parece que es más importante medir cuánto tiempo tardan las personas en salir del asombro del planteamiento.

La hipótesis que defienden Greene, Nystrom, Engell, Darley y Cohen (2004) y Bechara, Damasio, Damasio y Lee (1999) es que el juicio moral, en respuesta a violaciones familiares para nuestros antepasados primates, violaciones personales, es impulsado por respuestas socioemocionales; mientras que el juicio moral en respuesta a una violación moral distintivamente humana, impersonal, es o puede ser más *cognitivo*.

197

### III. 5. Soldado

En otro estudio sobre los criterios que emplean las personas para decidir a quién salvar en los dilemas de sacrificio, Bialek, Fugelsang y Friedman (2018) utilizaron a 558 participantes y encontraron que las preferencias pueden ser, al parecer, arbitrarias; su trabajo se basó en un escenario bélico hipotético sobre salvar la vida de tres soldados a costa de la vida de uno. Encontraron que la variable de la nacionalidad tenía mayor relevancia que otras variables, por ejemplo, edad; escogieron que el soldado sacrificado fuera de la misma nacionalidad que los soldados que se iban a salvar. También prefirieron que vivieran los soldados más jóvenes frente a los de mayor edad. Suponían que las personas preferirían el sacrificio de un soldado de una nacionalidad diferente de la de los soldados que se salvarían, porque se cree que hay una identificación de unidad nacional, especialmente tratándose de una guerra, pero en este caso, se trataba de una misión multinacional de la OTAN en la que hay varios países involucrados. Los autores (Bialek, Fugelsang y Friedman, 2018) concluyeron que los criterios para decidir quién debe ser sacrificado, en este tipo de dilemas, es una decisión multifactorial y por eso los participantes siempre quieren tener *más información*.

En el dilema del tranvía, muchos participantes quieren obtener más información sobre las personas atadas a las vías para poder tomar una

decisión (Swann Jr, Gómez, Dovidio, Hart y Jetten, 2010); por ejemplo, quieren saber cómo ha sido su comportamiento moral: si han sido *buenos* o *malos*. No así en el dilema de Sofía en el que prevalecen criterios altruistas o deontológico que suponen que la vida de uno de los hijos tiene el mismo valor que la del otro. De hecho, nadie se pregunta si uno de los hijos era, por ejemplo, más o menos obediente o cariñoso con su madre como criterio de sacrificio.

#### IV. Críticas a los dilemas

Como hemos vistos se considera la valía o no de los dilemas como una vía para descubrir las convicciones morales de las personas, cuando una simple variación de los escenarios o el modo en que estos se plantean, pueden hacer variar significativamente las respuestas dadas por los participantes. En el modo positivo o negativo en que se redacte el planteamiento del dilema se puede obtener una u otra respuesta: ¿Esto invalida su valor como instrumentos de investigación?

198 Petrinovich y O’neill (1996) consideran que una variación entre poner el énfasis en la palabra *salvar* o *matar* puede ser definitiva en los resultados que se obtengan, como un sesgo metodológico fuerte, aun cuando la consecuencia de la acción sea la misma. De hecho, los resultados que obtuvieron en sus estudios fueron que los participantes estaban más dispuestos a intervenir en *trolley*, cuando el énfasis estaba en el concepto de *salvar*, que cuando estaba en el concepto de *matar*. La explicación que ofrecen Petrinovich y O’neill es que un proceso adaptativo biológico y evolutivo ha predispuerto al ser humano intuitivamente a salvar a los demás por encima de querer matarlos.

En las múltiples versiones que se han hecho del dilema del tranvía algunas versiones se refieren a cinco hombres que trabajan en el mantenimiento de las vías (Greene, Nystrom, Engell, Darley y Cohen, 2004) y, en otras, a cinco personas atadas a las vías (Thomson, 1985); este tipo de enfoque, según Petrinovich y O’neill (1996) hace variar los resultados. Los participantes estarían más dispuestos a que mueran cinco *trabajadores*, pues podrían argumentar que *es parte de su riesgo laboral*, que cinco personas *atadas*, que son víctimas inocentes de un demente.

Algunos como Francis (2017) y Kittay y Carlson (2010) sugieren que existe una disociación entre el juicio moral y la acción moral, lo que significa que no necesariamente un juicio deba llevar a la realización correspondiente, o a la coherencia entre una creencia y el acto. En este sentido se puede argumentar que una creencia tiene distintos niveles de convicción y, sobre todo, de compromiso. El nivel de convicción podría determinar la realización o no de la acción.

Hasta ahora hay una amplia discusión sobre la utilidad de los dilemas morales hipotéticos (Christensen, Flexas, Calabrese, Gut, y Gomila, 2014) para obtener resultados fiables sobre la moral de los participantes. Sin embargo, otros estudios como los realizados por Christensen, Flexas, Calabrese, Gut, y Gomila (2014) arrojan resultados que consideran que *sí* es posible determinar las preferencias axiológicas y/o valorarles de las personas a través de los dilemas.

Por supuesto, *trolley* es un caso aparentemente inverosímil. El caso de Sofia, digamos *es diferente* dadas las múltiples historias del nazismo. Incluso se filmó una película en la que se recrea dicho dilema (Barish, Gerrity, Pakula y Starger, 1982). Sin embargo, como en otros, la principal crítica es que no reflejan escenas de la vida cotidiana y las decisiones que supuestamente se toman en el experimento, difícilmente se replicarían en la vida real. Sofia en el filme y en la novela finalmente comete suicidio por la culpa que jamás la abandona (Styron, 1979).

La críticas de Bauman, McGraw, Bartels y Warren (2014) coinciden en que los dilemas de sacrificio como *trolley* e incluso *Sofia*, no ofrecen evidencia valiosa, porque no representan escenarios posibles en la vida real. En cambio, cuando se procura abordar temas relevantes en los dilemas morales, como utilizar aviones no tripulados para actos terroristas, trabajo infantil, acoso sexual, discriminación, por mencionar algunos, los participantes se sienten incómodos y prefieren no participar, de ahí que *trolley* puede ser *más divertido*.

El realismo mundano, argumentan Bauman, McGraw, Bartels y Warren (2014), debería ser una característica deseable en todos los estudios experimentales sobre toma de decisiones morales. Utilizar escenarios altamente hipotéticos tienen poca o nula probabilidad de que se presenten en la vida y por tanto carecen de valor empírico.

Para Bauman, McGraw, Bartels y Warren (2014) un homicidio hipotético a nadie en el aula le importa. No despierta sentimientos de rechazo en el grupo, tampoco de empatía (Taylor, Eisenberg, Spinrad, Eggum y Sulik, 2013) o por lo menos, de enfrentamiento entre quienes creen que deberían desviar al tranvía y entre quienes no. Por el contrario, a algunos les parece *chistoso* lanzar al gordo.

Tampoco puede descartarse un argumento de irrelevancia psicológica porque, en efecto, si se toma en serio el problema dilemático documentado (Greene, Nystrom, Engell, Darley y Cohen, 2004), y cómo el cerebro puede recrear situaciones hipotéticas que cualitativamente se asemejan al proceso de toma de decisión en la vida real, se podría generar resultados conexos en la teoría general de la TD.

Según Greene, Nystrom, Engell, Darley y Cohen (2004), las implicaciones morales de la vida real y del laboratorio podrían ser las mismas. La toma de decisiones es un proceso altamente complejo en el que son muchos los factores que intervienen.

Aún con las críticas anteriores, consideramos el valor del estudio del dilema del tranvía y de Sofía porque siguen y seguirán atrayendo a los estudiosos de la moral y de las teorías de las decisiones. En el fondo, tratan de contrastar dos posturas filosóficas implícitas: utilitarismo contra deontologismo, actuar por los resultados o por principios que es un dilema al que muchos diariamente se enfrentan.

## V. Conclusiones

La forma en cómo tomamos decisiones sigue siendo problemática. No hay un manual de la razón práctica. Se puede computar algunos mecanismos como racionalización, emoción, marcadores somáticos, homeostasis o incluso, el absurdo, para caracterizar algunas acciones mentales y motoras. Sin embargo, lo cierto es que la TDM es multifactorial y difícilmente encasillada en una sola explicación positiva entre los *inputs* y los *ouputs*.

200

Consideramos que en el estudio de la TDM se deben cambiar viejos patrones o visiones que imposibilitan la entrada de innovadoras formas de apreciación conductual y de creencias. Para un proceso adaptativo e inteligente se requiere diferenciación de los escenarios cambiantes y poder de *inmersión* en esos contextos, conexos o inconexos, que esperan una respuesta inmediata, pero no por ello aleatoria.

La retroalimentación es tal vez la parte *más fina* de la TDM porque nos diferencia del resto de las especies. El humano es el único ser capaz de analizar, sintetizar y utilizar provechosamente la experiencia y obtención de resultados previos para la construcción de estrategias diferenciadas a futuro.

En cuanto a los correlatos neuronales de la decisión moral, no existe hasta ahora una teoría explicativa que logre demostrar que exista un área exclusivamente dedicada a los procesos de decisión. Si bien se considera que la vmPFC es la neuroanatomía más avanzada, no sería satisfactoria una hipótesis que la ubique como el lugar del *bien* y del *mal* en el hombre.

Por el contrario, existen más argumentos para creer que los procesos automáticos de decisión y toma de decisiones, anteriores a la conciencia, están mayormente presentes, aunque con ello se vulnera la agencia y la libre voluntad, que tanto nos gusta defender a los filósofos, especialmente iusnaturalistas.

Consideramos que las visiones deontológicas, utilitaristas, consecuencialista y emocionalistas a ultranza, son reduccionistas en el proceso más profundo y multidimensional que requiere la TDM.

Finalmente, la crítica a los dilemas morales, ya sean personales o impersonales para poder analizar y explicar el proceso de toma de decisión y su teoría limítrofe, también parecen insuficientes para explicar la complejidad de los fenómenos descritos. Los procesos adaptativos y cambiantes, y la posibilidad de adecuar el curso de la acción a cada circunstancia, nos debería convertir en *máquinas perfectas* de decisión; somos capaces de combinar la intuición, la sensibilidad y el razonamiento en cada caso único.

En la medida en que seamos conscientes de que están en curso procesos complejos, algunos controlables y otros no tanto, seremos los dueños de nuestro propio destino construido y el producto de las propias circunstancias.

### Referencias bibliográficas

- Armell, K. C., y Ramachandran, V. S. (2003). Projecting sensations to external objects: evidence from skin conductance response. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1523), 1499-1506. doi: 10.1098/rspb.2003.2364.
- Aulus Gellius. (1927). *Attic Nights, Volume I: Books 1-5. book 5, chapter 10*. (J. C. Rolfe, Trad.). Cambridge, EUA: Loeb Classical Library.
- Barish, K., Gerrity, W., Pakula, A., Starger, M. (productores) y Pakula, A. (director). (1982). *Sophie's choice* [Cinta cinematográfica]. Los Angeles, EUA: Universal Pictures.
- Bauman, C. W., McGraw, A. P., Bartels, D. M., y Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, 8(9), 536-554. doi: 10.1111/spc3.12131.
- Bechara, A., Damasio, H., Damasio, A. R., y Lee, G. P. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *JNeurosci. Journal of neuroscience*, 19(13), 5473-5481. doi: 10.1523/JNEUROSCI.19-13-05473.1999.
- Bernhard, R. M., Chaponis, J., Siburian, R., Gallagher, P., Ransohoff, K., Wikler, D. y Greene, J. D. (2016). Variation in the oxytocin receptor gene (OXTR) is associated with differences in moral judgment. *Social cognitive and affective neuroscience*, 11(12), 1872-1881. doi: 10.1093/scan/nsw103.
- Berthoud, H. R., Münzberg, H., y Morrison, C. D. (2017). Blaming the brain for obesity: integration of hedonic and homeostatic mechanisms. *Gastroenterology*, 152(7), 1728-1738. doi: 10.1053/j.gastro.2016.12.050.
- Bhakhthavatsalam, S., y Cartwright, N. (2017). What's so special about empirical adequacy? *European Journal for Philosophy of Science*, 7(3), 445-465. doi: 10.1007/s13194-017-0171-7.

- Bialek, M., Fugelsang, J., y Friedman, O. (2018). Choosing victims: Human fungibility in moral decision-making. *Judgment and Decision Making*, 13(5), 451-457. Recuperado de: <http://www.sjdm.org/journal/18/18612/jdm18612.pdf>.
- Bindra, D. (1978). How adaptive behavior is produced: a perceptual-motivational alternative to response reinforcements. *Behavioral and Brain Sciences*, 1(1), 41-52. doi: 10.1017/S0140525X00059380.
- Bland, A. R. (2012). Different varieties of uncertainty in human decision-making. *Frontiers in neuroscience*, 6, 85. doi: 10.3389/fnins.2012.00085.
- Buzzell, R. D., y Slater, C. C. (1962). Decision theory and marketing management. *Journal of Marketing*, 26(3), 7-16. doi: 10.2307/1248294.
- Christensen, J. F., Flexas, A., Calabrese, M., Gut, N. K., y Gomila, A. (2014). Moral judgment reloaded: a moral dilemma validation study. *Frontiers in psychology*, 5, 607. doi: 10.3389/fpsyg.2014.00607.
- Coutlee, C. G., y Huettel, S. A. (2012). The functional neuroanatomy of decision making: prefrontal control of thought and action. *Brain research*, 1428, 3-12. doi: 10.1016/j.brainres.2011.05.053.
- Damasio, A. (1996). *El error de Descartes*. Madrid, España: Destino.
- De la Fuente, R., y Leefmans, F. J. A. (2015). *Biología de la mente*. México D.F., México: Fondo de Cultura Económica.
- Doz, Y. L. (1996). The evolution of cooperation in strategic alliances: initial conditions or learning processes? *Strategic management journal*, 17(S1), 55-83. doi: 10.1002/smj.4250171006.
- Drevets, W. C., Price, J. L., y Furey, M. L. (2008). Brain structural and functional abnormalities in mood disorders: implications for neurocircuitry models of depression. *Brain structure and function*, 213(1-2), 93-118. doi: 10.1007/s00429-008-0189-x.
- Evans, A. M., Dillon, K. D., y Rand, D. G. (2014). Fast but Not Intuitive, Slow but Not Reflective: Decision Conflict Drives Reaction Times in Social Dilemmas. *Journal of Experimental Psychology: General*, 144(5), 951-966. doi: 10.2139/ssrn.2436750.
- Evans, J. S. B., y Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, 8(3), 223-241. doi: 10.1177/1745691612460685.
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, (5), 5-15.
- Francis, K. B. (2017). *Moral cognition: An interdisciplinary investigation of judgment versus action*. (tesis doctoral) Recuperado de: <https://pearl.plymouth.ac.uk/handle/10026.1/9583>.
- Gazzaniga, M. S. (2012). *¿Quién manda aquí? El libre albedrío y la ciencia del cerebro*. Barcelona, España: Paidós.
- Gintis, H. (2007). A framework for the unification of the behavioral sciences. *Behavioral and brain sciences*, 30(1), 1-16. doi: 10.1017/S0140525X07000581.

- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., y Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44(2), 389-400. doi: 10.1016/j.neuron.2004.09.027.
- Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological review*, 108(4), 814. doi: 10.1037/0033-295x.108.4.814.
- Hall, D., Dhillia, A., Charalambous, A., Gogos, J. A., y Karayiorgou, M. (2003). Sequence variants of the brain-derived neurotrophic factor (BDNF) gene are strongly associated with obsessive-compulsive disorder. *The American Journal of Human Genetics*, 73(2), 370-376. doi: 10.1086/377003.
- Higgs, S. (2016). Cognitive processing of food rewards. *Appetite*, 104, 10-17. doi : 10.1016/j.appet.2015.10.003.
- Hinton, G. E., y Becker, S. (1990, January). An unsupervised learning procedure that discovers surfaces in random-dot stereograms. *Proceedings of the International Joint Conference on Neural Networks*, 1, 218-222. Recuperado de: <http://www.cs.utoronto.ca/~hinton/absps/beckejcnn.pdf>.
- Hume, D. (2005). *Tratado de la Naturaleza Humana*. Madrid, España: Tecnos.
- Kamm, F. M. (2015). *The trolley problem mysteries*. Oxford, Inglaterra: Oxford University Press.
- Kant, I. (1993). Sobre un presunto derecho a mentir por filantropía. En R. Rodríguez Aramayo (Ed.) y J. M. Palacios (Trad.), *Teoría y Práctica* (pp. 423-430). Madrid, España: Tecnos.
- Kant, I. (2018). *Fundamentación de la metafísica de las costumbres*. CDMX, México: Ariel.
- Kaplan, R. S., y Mikes, A. (2012). Managing risks: a new framework. *Harvard Business Review*, 90(6), 48-60. Recuperado de: <https://hbr.org/2012/06/managing-risks-a-new-framework>.
- Katz, D. (1964). The motivational basis of organizational behavior. *Behavioral science*, 9(2), 131-146. doi: 10.1002/bs.3830090206.
- Kittay, E. F., y Carlson, L. (Eds.). (2010). *Cognitive disability and its challenge to moral philosophy*. Chichester, Inglaterra: John Wiley & Sons.
- Kohlberg, L. (1969). Stage and sequence: The cognitive development approach to socialization. En D. A. Goslin (Ed.), *Handbook of socialization theory* (pp. 347-480). Chicago, EUA: McNally.
- Kornmeier, J., y Bach, M. (2012). Ambiguous figures—what happens in the brain when perception changes but not the stimulus. *Frontiers in human neuroscience*, 6, 51. doi: 10.3389/fnhum.2012.00051.
- Lanteri, A., Chelini, C., y Rizzello, S. (2008). An experimental investigation of emotions and reasoning in the trolley problem. *Journal of Business Ethics*, 83(4), 789-804. doi: 10.1007/s10551-008-9665-8.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and brain sciences*, 8(4), 529-539. doi: 10.1017/S0140525X00044903.

- MacKinnon, D. P., Fairchild, A. J., y Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, 58, 593-614. doi: 10.1146/annurev.psych.58.110405.085542.
- Maslow, A. H. (1943). A theory of human motivation. *Psychological review*, 50(4), 370-396.
- Mill, J. S. (1980). *El utilitarismo*. Buenos Aires, Argentina: Aguilar.
- Morin, E., y Pakman, M. (1994). *Introducción al pensamiento complejo*. Barcelona, España: Gedisa.
- Northoff, G., y Bermpohl, F. (2004). Cortical midline structures and the self. *Trends in cognitive sciences*, 8(3), 102-107. doi: 10.1016/j.tics.2004.01.004.
- Papageorgiou, G. K., Sallet, J., Wittmann, M. K., Chau, B. K., Schüffelgen, U., Buckley, M. J., y Rushworth, M. F. (2017). Inverted activity patterns in ventromedial prefrontal cortex during value-guided decision-making in a less-is-more task. *Nature communications*, 8, 1886. doi: 10.1038/s41467-017-01833-5.
- Pereboom, D. (2014). *Free will, agency, and meaning in life*. Oxford, Inglaterra: Oxford University Press.
- Petrinovich, L., y O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology*, 17(3), 145-171. doi: 10.1016/0162-3095(96)00041-6.
- Platón. (2013). *El Fedón*. Barcelona, España: Minimal.
- Ridderinkhof, K. R., Van Den Wildenberg, W. P., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and cognition*, 56(2), 129-140. doi: 10.1016/j.bandc.2004.09.016.
- Rosas, A., Alfonso, E. C., Arciniegas, M. A., y Gómez, A. A. (2013). ¿Decisión Utilitarista O Decisión Aleatoria?: Crítica a una tesis atrincherada en la neurociencia cognitiva. *Ideas y valores: Revista Colombiana de Filosofía*, 62(153), 179-199.
- Rosas, A., Arciniegas, A., Cavides, E., y Arciniegas, A. (2014). La neuropsicología del juicio moral. Sobre las causas de respuestas contraintuitivas a los dilemas morales. *Praxis filosófica*, (38), 89-106. doi: 10.25100/pfilosofica.v0i38.3526.
- Singer, T., y Tusche, A. (2014). Understanding others: Brain mechanisms of theory of mind and empathy. *Neuroeconomics*, 513-532. doi: 10.1016/B978-0-12-374176-9.00017-8.
- Stajkovic, A. D., & Luthans, F. (1998). Social cognitive theory and self-efficacy: Goin beyond traditional motivational and behavioral approaches. *Organizational dynamics*, 26(4), 62-74. doi: 10.1016/S0090-2616(98)90006-7.
- Styron, W. (1979). *Sophie's Choice*. New York, EUA: Random House.
- Swann Jr, W. B., Gómez, Á., Dovidio, J. F., Hart, S., y Jetten, J. (2010). Dying and killing for one's group: Identity fusion moderates responses to intergroup versions of the trolley problem. *Psychological Science*, 21(8), 1176-1183. doi: 10.1177/0956797610376656.

- Taylor, Z. E., Eisenberg, N., Spinrad, T. L., Eggum, N. D., y Sulik, M. J. (2013). The relations of ego-resiliency and emotion socialization to the development of empathy and prosocial behavior across early childhood. *Emotion, 13*(5), 822-831. doi: 10.1037/a0032894.
- Thomson, J. (1985). The Trolley Problem. *The Yale Law Journal, 94*(6), 1395-1415.
- Van den Bos, W., y Güroğlu, B. (2009). The role of the ventral medial prefrontal cortex in social decision making. *Journal of Neuroscience, 29*(24), 7631-7632. doi: 10.1523/JNEUROSCI.1821-09.2009.
- Verplaetse, J. (2014). *Moral brain. Essays on the Evolutionary and Neuroscientific Aspects of Morality*. Luxemburgo, Luxemburgo: Springer.
- Warren, S. J., Dondlinger, M. J., McLeod, J., y Bigenho, C. (2012). Opening The Door: An evaluation of the efficacy of a problem-based learning game. *Computers & Education, 58*(1), 397-412. doi: 10.1016/j.compedu.2011.08.012.
- Wisner, B. L., Jones, B., & Gwin, D. (2010). School-based meditation practices for adolescents: A resource for strengthening self-regulation, emotional coping, and self-esteem. *Children & Schools, 32*(3), 150-159. doi: 10.1093/cs/32.3.150.

