Mº Natalia Zavadivker [1]

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

"No es lo que conoces, sino lo que ponderas". Una propuesta alternativa sobre las bases neuronales y cognitivas implicadas en la "suerte moral".

"It's not what you know, but what you weight". An alternative offer about the neural and cognitive basis implied in "moral luck".

"Não é o que conheces, mas sim o que ponderas". Uma proposta alternativa sobre as bases neuronais e cognitivas implicadas na "sorte moral".

[1] Dra. en Filosofía. Instituto de Biotecnología- Fac. de Bioquímica, Química y Farmacia- Universidad Nacional de Tucumán (UNT)- Ayacucho 471 (4000) San Miguel de Tucumán. Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET)- Av. Rivadavia 1917- CABA. La Plata 1034 (4000) San Miguel de Tucumán- e-mail: zavadivker@yahoo.com.ar.

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

Resumen

Se analiza críticamente la hipótesis de Young et. al. según la cual los únicos factores moralmente evaluables en casos de daños no intencionales desencadenantes de resultados negativos ('suerte moral'), son de índole cognitiva (valor de verdad y justificación de las creencias del agente responsable). Como objetivo específico se propone un enfoque superador sobre los posibles criterios intervinientes en las evaluaciones morales que contemple el influjo de factores pragmáticos (motivaciones, impulsos, comportamientos automáticos o ritualizados, ponderación de valores, etc.). El propósito más amplio es cuestionar el supuesto de que es posible derivar un juicio moral de la mera interpretación de los estados mentales del agente, sin la mediación de una tarea cognitiva "extra" consistente en asignar un valor o estimación moral a tal contenido. Se proponen algunas variantes metodológicas de los experimentos analizados que podrían contribuir a detectar si las personas apelan a factores no cognitivos (como la ponderación de los valores del agente en comparación con los propios) cuando juzgan acciones ligadas a la 'suerte moral'. También se proponen experimentos que contribuirían a detectar la posible activación de regiones cerebrales específicas presuntamente intervinientes en la tarea de valorar estados mentales ajenos, bajo el supuesto de que tal operación no es identificable con (ni reductible a) la mera lectura de mente.

Palabras clave: suerte moral; evaluación moral; creencias / razones; resultados / intenciones; cognitivismo vs. pragmatismo; ponderación de valores; lectura de mente; interpretación vs. valoración.

Abstract

We critically analyze the hypothesis by Young et. al. according to which the only factors morally evaluables in cases of unintentional damage that result in negative outcomes ('moral luck') are purely cognitives (false beliefs and bad reasons of responsible agent). Since specific aim proposes a more comprehensive approach in relation to possible criteria involved in moral evaluations that addresses the possible influence of pragmatic factors (motivations, impulses, automatic or ritualized behaviors, weighting values, etc.). A broader aim is to question the assumption that it's possible to derive a moral judgment of mere interpretation of mental states (beliefs and intentions) of an agent, without the mediation of an "extra" cognitive task: assigning a value or moral estimate to the mental content. We propose some methodological variants of analized experiments that could contribute, first, to detect if people appeal to non-cognitive factors (such as weighting's values of the agent compared to itself) when judging moral actions linked to 'moral luck'; and second; to detect activation of specific brain regions involved in the task of assessing mental states of others, under the assumption that this task is not identifiable with (or reducible to) the mere reading of mind.

Key words: moral luck; moral evaluate; beliefs/reasons; outcomes/intentions; cognitivism vs. pragmatism; weighting values; mind reading; interpretation vs. assessment.

Resumo

Se analisa criticamente a hipótese de Young et. al. segundo a qual os únicos fatores moralmente avaliáveis nos casos de danos não intencionais desencadeantes de resultados negativos ("sorte moral"), são de índole cognitiva (valor de verdade e justificação das crenças do agente responsável). Como objetivo específico propõe-se um enfoque superador sobre os possíveis critérios intervenientes nas avaliações morais que contemple o influxo de fatores pragmáticos (motivações, impulsos, comportamentos automáticos ou ritualizados, ponderação de valores, etc.). O propósito mais amplo é questionar o suposto de que é possível derivar um juízo moral da mera interpretação dos estados mentais do agente, sem a mediação de uma tarefa cognitiva "extra" consistente em atribuir um valor ou estimação moral a tal conteúdo. Propõe-se algumas variantes metodológicas dos experimentos analisados que poderiam contribuir a detectar se as pessoas apelam a fatores não cognitivos (como a ponderação dos valores do agente em comparação com os próprios) quando julgam ações ligadas a 'sorte moral'. Também propõe-se experimentos que contribuiriam para detectar a possível ativação de regiões cerebrais específicas supostamente intervenientes na tarefa de avaliar estados mentais alheios, sobre o suposto de que tal operação não é identificável com (nem redutível a) mera leitura da mente.

Palavras chaves: sorte moral; avaliação moral; crenças/ razões; resultados/intenções; cognitivismo vs. pragmatismo; ponderação de valores; leitura da mente; interpretação vs. avaliação.

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

Este artículo se propone analizar críticamente la hipótesis sustentada y constrastada experimentalmente por Young, Nichols y Saxe en su artículo "Investigating the Neural and Cognitive Basis of Moral Luck: It's Not What You Do but What You Know" (2010); a los fines de sugerir una explicación alternativa y complementaria a la sustentada por los autores en relación a los factores implicados en la evaluación moral de comportamientos asociados a daños accidentales que redundan en resultados negativos, aun cuando el agente no tenía la intención de provocarlos. Dicha hipótesis alternativa está a su vez destinada a cumplir dos objetivos de alcance más amplio:

1º) Proponer un enfoque superador en relación a los posibles criterios que intervienen en las evaluaciones morales, que no se limite a los aspectos meramente cognitivos-tales como el valor de verdad o la justificación de las creencias- sino que contemple el posible influjo de factores pragmáticos (motivaciones, impulsos, comportamientos automáticos o ritualizados, ponderación de valores, etc.).

2º) Cuestionar el supuesto según el cual es posible derivar un juicio moral de la mera interpretación de los estados mentales (creencias e intenciones) de un agente, sin la mediación de una operación o tarea cognitiva "extra" consistente en asignar un valor o estimación moral a dicho contenido mental. El supuesto de base es que, si efectivamente el acto de valorar intenciones, creencias, situaciones, comportamientos, etc. (ya sea en términos morales o no) implica una tarea cognitiva per se, no reductible a la mera tarea de "lectura de mentes", entonces debería haber alguna región o regiones cerebrales reclutadas para el cumplimiento de dicha función.

Las explicaciones filosóficas tradicionales (Nagel, 1979, Williams, 1982) sugieren que las personas tendemos a juzgar los actos por sus resultados, adjudicando culpa moral a aquellos individuos cuyas acciones condujeron a desenlaces desafortunados, pero aprobando, no condenando o haciéndolo en menor medida si las mismas acciones hubieran dado lugar a resultados afortunados o neutrales. En otras palabras, nuestros juicios morales estarían directamente influidos por los resultados de las acciones, de modo tal que los agentes desafortunados serían más condenados moralmente que aquellos cuyas acciones redundaron en resultados positivos (como asumir el riesgo de probar los efectos de una droga medicinal desconocida, y obtener una mejoría significativa, lo que nos llevaría a aprobar, e incluso ponderar positivamente dicha acción), neutrales (como correr una picada automovilística en una zona concurrida y que todos salgan ilesos, lo que podría llevarnos a no condenar el hecho) o no tan malos (por ej., que se derrumbe un edificio mal construido cuando no había nadie adentro, lo que probablemente llevaría a una condena menor que si el hecho hubiera dejado un saldo de varios muertos y heridos). Ahora bien, dicha propensión a juzgar las acciones por sus resultados es considerada en la literatura psicológica como un sesgo irracional a-posteriori, asociado a la falacia consistente en tratar los eventos pretéritos como predecibles (una vez que ocurre el mal resultado, solemos creer que sabíamos que ocurriría). Para algunos autores (Greene et al. 2001, 2004; Haidt, 2001) la constatación de resultados desafortunados (por ej., la observación del cadá-

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

ver de un niño muerto en un accidente) desencadena respuestas emocionales que instan al observador a culpar moralmente al agente responsable (por ej., el padre que lo atropelló accidentalmente), independientemente de la estimación de las creencias e intencionalidad de dicho agente. Young et. al. (2010a) aducen que el factor "suerte" no debería ser lo que determine la racionalidad de una acción. De acuerdo a estos autores, para que nuestras evaluaciones estén racionalmente fundadas, deberíamos tener en cuenta las razones de quien ejecuta la acción para prever, en un contexto de incertidumbre, las posibles consecuencias de sus actos, apoyadas a su vez en el valor de verdad de sus creencias (vale decir, sus presunciones previas acerca de cuán exitosas o desafortunadas pueden llegar a ser las consecuencias de sus acciones).

Young et al. cuestionan la hipótesis según la cual las evaluaciones morales de las personas se basan exclusivamente en la consideración de los resultados. Proponen testear la hipótesis alternativa de que estos juicios incluyen también la evaluación de las creencias (verdaderas o falsas) del agente, así como la pertinencia de las razones en que éstas se apoyan. Suponen que el agente que juzga atribuirá la mala suerte de un sujeto más a sus falsas creencias (basadas además en malas razones) que a los resultados desafortunados como tales, de modo tal que, por ej., quienes basen sus actos en creencias falsas sobre sus presuntas consecuencias, serán juzgados como moralmente culpables aun cuando hayan tenido la buena suerte de que el resultado negativo no ocurra. De este modo, Young et al. pretenden que su estudio preste soporte a una explicación racionalista de la suerte moral: las asimetrías a la hora de juzgar al agente no estarían motivadas primariamente por los sesgos en el resultado, sino por una estimación del valor de verdad de sus creencias y la pertinencia de sus razones (por ej., si los agentes están justificados al pensar que sus actos no causarán daños a terceros). Argumentan que aun los niños de tres años son capaces de distinguir (y formular diferentes juicios morales y sociales) entre eventos incontrolables debidos exclusivamente al azar (por ej.: que llueva en un partido de soccer), y eventos desafortunados que dependen de la acción u omisión de algún sujeto.

Para testear su hipótesis, Young et al. presentan a un grupo de sujetos experimentales el siguiente escenario moral: "Mitch prepara un baño para su hijo de dos años, que está parado junto a la bañera, cuando suena el teléfono en la habitación contigua. Mitch le dice a su hijo que permanezca parado, creyendo que éste le obedecerá, y sale de la habitación por un momento".

Los posibles resultados del comportamiento de Mitch son:

- Que al regresar el niño esté parado ileso junto a la tina (resultado neutral).
- Que al regresar el niño esté flotando ahogado en la bañera (resultado desafortunado).

A pesar de que a primera vista todo da a entender que la condena moral a Mitch dependerá crucialmente del resultado de sus acciones (vale decir, de si el niño se ahoga en la bañera o permanece ileso), Young et. al. procuran testear si la evaluación moral de potenciales observadores toma en cuenta aspectos tales como las creencias (verdaderas o falsas) de Mitch acerca de si el niño se quedará o no quieto ante sus órdenes, y sus razones (buenas o malas) para sustentar tales creencias. Sus predicciones son las siguientes:

- Los observadores juzgarán las falsas creencias como menos justificadas que las creencias verdaderas (dado que la creencia "el niño no se moverá" del agente desa-

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

fortunado es falsa, los observadores la percibirán como menos justificada que la del agente afortunado).
- Los observadores asignarán mayor culpa moral al agente que actúan sobre falsas creencias que al que actúa sobre creencias verdaderas, aun cuando éstas redunden en los mismos resultados neutrales.
- A la hora de juzgar moralmente, los observadores no sólo tomarán en cuenta el valor de verdad de la creencia, sino también las razones que permiten justificarla.
- Si el comportamiento del agente moral que sustenta una falsa creencia redunda en un desenlace neutro (no sucede el hecho desafortunado), los observadores juzgarán dicho resultado como producto exclusivo de la buena suerte. En otras palabras, el resultado neutro no afectará el valor de verdad de la creencia: si ésta era falsa lo seguirá siendo con independencia del resultado.

La principal diferencia entre juzgar un acto por sus resultados y juzgarlo basándose en las creencias y justificaciones del agente, es que si no hiciéramos esto último deberíamos considerar un resultado desafortunado como producto exclusivo del azar (factor que no puede ser controlado y, por ende, exime al agente de toda responsabilidad moral). Sin embargo, ante los eventos catastróficos lo que solemos hacer es evaluar los posibles factores causales (concretamente, el error humano que dio lugar al desenlace infortunado). Young et al. conjeturan que lo que juzgarían los observadores en tales casos son las creencias del agente, y en qué medida éstas están justificadas. Admiten que tal justificación suele darse a posteriori de la ocurrencia del hecho, con lo cual el sesgo asociado a los resultados sigue operando en gran medida: percibimos cierto desenlace como obvio y esperable una vez que el hecho se produjo. Sin embargo, eso no significa que el observador no efectúe conjeturas racionales ligadas a las posibles causas del desenlace fatal: lo que los observadores inferirían (y, por ende, evaluarían moralmente), es que el agente responsable sostenía una creencia errónea sobre los posibles efectos de su accionar, y ésta estaba además débil o inadecuadamente justificada. Young et al. aclaran además que no abordarán aspectos normativos del juicio moral (por ej.: si sería correcto o no culpar a un agente por sustentar una creencia falsa). Simplemente investigarán si los agentes desafortunados suelen ser juzgados de hecho como más culpables porque sus falsas creencias son percibidas como menos justificadas.

A los fines de testear por separado las falsas creencias y los resultados negativos, Young et. al. introdujeron una nueva variable en el escenario moral testeado: la del agente 'extra-afortunado' (extra-lucky). Éste tiene las mismas creencias que el agente desafortunado ("Mi hijo me obedecerá y se quedará quieto"), pero, por un golpe de suerte extraordinario, su comportamiento no conduce al resultado negativo (su hijo no se ahoga). El objetivo era comprobar si los observadores condenarían moralmente al agente 'extra-afortunado' con el mismo rigor con que juzgaron al agente desafortunado. Conjeturaron que Mitch sería juzgado como moralmente culpable por sustentar una creencia falsa, aun en ausencia del resultado negativo.

Los autores presentaron a los participantes 54 escenarios morales, con 9 variaciones por cada uno. En todas las condiciones las acciones (dejar al niño sólo frente a la tina) y creencias del agente ("el niño no se moverá") eran idénticas, lo que variaba eran las razones esgrimidas para justificarlas, que podían ser

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

'buenas' ("El hijo de Mitch siempre le obedece, por lo tanto, Mitch cree que su hijo lo esperará por un momento"), 'malas' ("El hijo de Mitch nunca le obedece, pero Mitch cree que su hijo lo esperará por un momento"), o 'inespecíficas' (no se explicitan las razones del agente). Los resultados podían ser neutrales (el niño no se ahoga) o malos (el niño se ahoga), y las creencias del padre podían ser verdaderas o falsas. Esto daba lugar a tres situaciones posibles:

1º) Creencia verdadera ("Mi hijo no se moverá") apoyada en buenas razones ("Siempre hace lo que le digo"): resultado neutro (el niño resulta ileso).
2º) Creencia falsa ("Mi hijo no se moverá") apoyada en malas razones [1] ("Nunca hace lo que le digo"): resultado negativo (el niño se ahoga).
3º) Creencia falsa ("Mi hijo no se moverá"), razones no especificadas: resultado neutro (tal es el caso del agente 'extra-afortunado', quien por un golpe de suerte no obtiene consecuencias negativas pese a sus falsas creencias).

Los observadores debían juzgar, en una escala del 1 al 7, cuán razonables o justificadas eran las creencias del agente (experimento conductual 1), y en qué grado éste era culpable de sus actos (experimento conductual 2), a los fines de detectar posibles correlaciones entre la pertinencia de las razones del agente y la asignación de culpa moral.

Los experimentos conductuales fueron a su vez complementados con estudios basados en técnicas de RMNf (resonancia magnética nuclear funcional), a los fines de testear si la actividad en las regiones cerebrales implicadas en el razonamiento moral "distinguían" entre creencias verdaderas y falsas y entre buenas o malas razones. La conjetura de base era que, si la verdad y la justificación son propiedades relevantes en la producción de juicios morales, éstas podrían generar activaciones diferenciales durante la evaluación de las creencias del agente, aun desconociendo los resultados de la acción. En contraste, también esperaban observar si esas activaciones (que indicarían que los sujetos están atendiendo al contenido y justificación de las creencias) se daban con mayor intensidad una vez que los sujetos conocían los resultados (positivos o negativos) de la acción del agente. Los estímulos y la presentación fueron idénticos a los del experimento conductual, con la excepción de que los participantes sólo respondieron la pregunta relativa al grado de culpabilidad de Mitch, de acuerdo a una escala de 4 puntos (1: "no del todo culpable"; 4: "muy culpable"). A fin de identificar en cada sujeto individual las regiones cerebrales relevantes (ROI) para la atribución de estados mentales -unión temporo-parietal derecha e izquierda (UTPD y UTPI, en inglés RTPJ y LTPJ) áreas implicadas en la "lectura de mente"-razonamiento acerca de estados mentales-; precuneus (PC) y corteza prefrontal medial ventral y dorsal (CPFM) asociada a la comprensión del comportamiento de otras personas en términos de sus creencias e intenciones- los sujetos participaron en la misma sesión de escaneo de cuatro tareas de razonamiento acerca de estados mentales, las cuales fueron contrastadas con otras tareas asociadas a representaciones físicas (mapas, señales, fotografías, etc.).

Para obtener los resultados, se analizaron los efectos de las razones del agente sobre: las creencias (verdaderas, falsas), el resultado de la acción (neutral, malo), los juicios de los participantes sobre la culpabilidad moral del agente, la justificación de las creencias, y las respuestas neurales en cada región de interés (ROI).

[1] No me queda claro por qué los autores decidieron no especificar las razones del agente 'extra-afortunado' para sustentar sus creencias, cuando ellos mismos conjeturaron que los observadores tendrán en cuenta, a la hora de emitir un juicio moral, tanto las creencias del agente como las razones en que éstas se basan.

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

He aquí algunos de los resultados más relevantes:

- En primer término, y parcialmente en contra de la hipótesis sustentada, los investigadores pudieron comprobar que los malos resultados ejercen un sesgo motivacional fundamental a la hora de incitar a los observadores a examinar los estados mentales del agente y juzgar tanto el valor de verdad de sus creencias, como su justificación. Proponen un modelo de inputs cognitivos en el que diversos mecanismos se retroalimentan mutuamente: los resultados (malos o neutrales) influyen directamente sobre la asignación de culpa moral al agente, lo que incita a los observadores a formular juicios sobre cuán justificadas son sus creencias (juicios de justificación de las creencias). A su vez, dichos juicios estarían influidos por la verdad o falsedad de las creencias del agente, y por las razones en que se apoyan (buenas, malas o inespecíficas). Al respecto, Young et. al. pudieron comprobar que el agente fue culpado por sustentar una falsa creencia, con independencia de las razones a favor de la misma, lo que estaría indicando que la verdad o falsedad de las creencias parece ejercer un peso importante sobre su justificación: si la creencia es verdadera, tenderemos a pensar que está fundada en buenas razones; mientras que si es falsa supondremos que se funda en malas o débiles razones. Algunos autores (Royzman y Kumar, 2004) interpretan esta tendencia como un prejuicio irracional; mientras que para otros (Richards, 1986, Rosebury, 1995) se trataría de una operación cognitiva normativamente legítima.

En síntesis, en consonancia con las aproximaciones tradicionales no racionalistas sobre la suerte moral, los juicios sobre la culpabilidad moral del agente fueron afectados directamente por los resultados, de modo tal que los desenlaces desafortunados llevaban a los observadores a juzgar los estados mentales del agente y examinar cuán justificadas eran sus creencias. Se observó una aparente interdependencia entre varios factores: tanto los juicios morales como los juicios de justificación de las creencias fueron influidos por los malos resultados particularmente cuando los agentes tenían malas razones para sustentar sus creencias (por ej., cuando Mitch no tenía razones de peso para creer que su hijo se quedaría quieto -antes bien, había importantes razones para pensar lo contrario- y además ocurrió el resultado negativo-muerte del niño-, fue juzgado más severamente que cuando ocurrió el resultado negativo, pero Mitch tenía buenas razones para creer que el niño permanecería quieto). Pero, al mismo tiempo, las mismas falsas creencias basadas en las mismas malas razones fueron juzgadas más severamente cuando conducían al resultado negativo que cuando daban lugar a resultados neutrales. Dada esta doble influencia sobre los juicios morales (consideración de los resultados y consideración de la justificación de las creencias), Young et. al. se preguntaron si los participantes juzgaron a los agentes que causaron malos resultados porque primero juzgaron sus creencias como menos justificadas, o si, a la inversa, fueron los malos resultados los que los llevaron a juzgar las creencias de los agentes. Proponen una explicación aparentemente compatible con el fenómeno de la "perplejidad moral" descripto por Haidt (2001): los participantes inicialmente formulan juicios de desaprobación moral instantáneos basándose en los malos resultados, en la medida en que su carácter catastrófico despierta reacciones emocionales inmediatas que incitan a asignar culpabilidad al agente que los ocasionó. Luego buscan justificar racionalmente sus juicios evaluando cuán fundamentadas eran creencias del agente, pero se trata de una justificación post-hoc, dado que el examen de las razones del agente no fue lo que causó el juicio moral, sino que éste se disparó espontáneamente como resultado de la activación de mecanismos emocionales.

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

Los juicios de justificación de las creencias aparecen sólo cuando los participantes son instados a dar razones de sus evaluaciones morales.

Los investigadores también relacionan sus resultados con otro sesgo conocido como "efecto Knobe": cuando alguien causa un mal resultado por efecto de la imprudencia, desidia o falta de recaudos, los observadores tienden a adjudicar al agente cierta intencionalidad; lo que no sucede cuando la misma acción da lugar a resultados neutrales u ocasionalmente positivos. Dicho efecto también sería aplicable a los estados epistémicos, como la posesión de creencias acerca de sucesos futuros. Así, por ej., si Mitch causó un mal resultado (su hijo yace ahogado en la bañera) los observadores tenderán a pensar que sabía lo que iba a suceder, con lo cual su comportamiento tiende a interpretarse como parcialmente intencional. Algo similar sucede cuando un empresario genera un daño accidental en el medioambiente como consecuencia de la implementación de una actividad económica rentable: la gente tiende a pensar que el agente conocía los efectos de dicha actividad, mientras que no asignamos el mismo estado mental cuando el resultado negativo no ocurre.

En cuanto a los resultados obtenidos por RMNf, la expectativa de Young et. al. era que, si los malos resultados conducen a juicios morales más severos, lo que induce a los observadores a examinar los estados mentales (en este caso las creencias) del agente; cabe esperar que los resultados negativos se correlacionen con una activación aumentada de regiones cerebrales asociadas al razonamiento acerca de creencias. Efectivamente, encontraron una respuesta robusta en las dos regiones asociadas al razonamiento sobre creencias en contextos morales y no morales (UTPD y UTPI), ante la presentación del contenido de las creencias y razones del agente. Estas regiones presentaron una importante activación mientras los participantes leían acerca de las creencias, con independencia de si éstas eran justificadas o injustificadas. Luego, mientras los participantes emitían sus juicios morales (lo que no sucedió en igual medida cuando las creencias fueron presentadas) se observó una significativa respuesta en la UTPD y la UTPI para los malos resultados, no así para los resultados neutrales. Esto sugirió que fue efectivamente la influencia de los malos resultados (y no la de las falsas creencias) lo que instó a los observadores a examinar las creencias y razones de los agentes, en contra de lo que éstos formularon en sus juicios explícitos (en los que ponían énfasis en las malas razones de los agentes, independientemente de los resultados). En otras palabras, la respuesta neural ante las malas razones del agente fue mayor cuando los resultados fueron negativos. También hallaron un efecto no esperado sobre la verdad sólo en la UTPD: hubo una respuesta mayor en esta región cuando las creencias del agente eran verdaderas que cuando eran falsas. Sin embargo, cuando los participantes emitían sus juicios morales, el cambio en la señal percentil promedio en la UTPD y la UTPI no fue afectado por las características de las creencias (verdad y justificación). De acuerdo a Young et. al., estos resultados admiten dos interpretaciones: o bien esta región está asociada a la capacidad de forjarse representaciones del contenido de las creencias, pero no de su verdad o justificación; o bien estas regiones pueden contener información sobre la verdad o la justificación de las creencias, pero la misma está asociada a una activación por debajo del umbral detectable por el aparato de medición.

El objetivo de este trabajo es, en primer lugar, proponer una hipótesis alternativa y complementaria a la explicación de Young et. al. sobre los criterios que tendrían en cuenta los observadores a la hora de juzgar acciones no intencionales, pero que pueden desencadenar resultados negativos. Un segundo objetivo, de alcance más general y profundo, es especular acerca

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

de si los autores están tomando en consideración todos los mecanismos psicológicos necesarios para la generación de juicios morales, ya que éstos parecen asumir que basta con la interpretación del contenido de los estados mentales del agente (creencias e intenciones), más la consideración del valor de verdad y la justificación de sus creencias, como requisitos necesarios y suficientes para la evaluación moral. A mi modo de ver, el acto mismo de valorar (evaluar moralmente) es una operación mental en sí misma, separable de la mera atribución de creencias e intenciones (una suerte de operación meta-cognitiva que nos lleva a adjudicar un peso o valor moral determinado al contenido mental del agente, y que no puede agotarse, o no es reductible, a la mera interpretación de sus estados mentales).

Comenzaré con una breve referencia a la línea de investigación seguida por Saxe, Young y su equipo, a los fines de desentrañar los supuestos en los que, a mi juicio, fundan sus hipótesis, sobre todo en lo atinente a las bases neurales implicadas en el razonamiento moral. Este equipo de investigadores realizó varios estudios previos que dan cuenta de la activación selectiva de la Unión Temporoparietal Derecha (UTPD) cuando los sujetos experimentales procesan información relativa a los estados mentales–creencias e intenciones- de otras personas (Young, Cushman, Hauser, Saxe 2007, Young y Saxe 2008, Young y Saxe 2009a. y 2009b, Young, Dodell-Feder y Saxe 2010). De allí que dicha localización cerebral haya sido asociada a la "teoría de la mente" o capacidad de "lectura de mentes". Dado que la mayor o menor actividad de la UTPD parece correlacionarse con una mayor o menor facilidad para interpretar los estados mentales de otras personas; cuanto menor es la actividad en dicha región, nuestros juicios morales tienden a ser más utilitaristas (vale decir, más orientados a la evaluación de los resultados, sin tener en cuenta, o considerando en menor medida, las creencias e intenciones del agente); mientras que, a mayor actividad de la UTPD, mayor es la tendencia a evaluar el comportamiento del agente en base a sus intenciones. Para corroborar si nuestras evaluaciones morales consideran en mayor medida la intencionalidad del agente o los resultados de la acción, los investigadores recurrieron típicamente al siguiente escenario: una joven llamada Grace procura, en una primera versión, asesinar a su amiga envenenando su café con un polvillo blanco similar en apariencia al azúcar, pero se equivoca y coloca azúcar común en lugar del veneno, con lo cual el resultado negativo (muerte de la amiga) no ocurre. En la segunda versión, Grace no tiene ninguna intención de envenenar a su compañera, pero por error endulza su café con veneno -creyendo que se trata de azúcar-, y su amiga muere. Tenemos entonces dos escenarios principales:

1º) Intención de provocar daño (asesinar a la colega) + falsa creencia sobre los medios adecuados para alcanzar el fin ('el polvo blanco es veneno') = resultado neutro (a la amiga de Grace no le ocurre nada).

2º) No hay intención de provocar daño (Grace sólo quiere llevarle el café a su colega) + falsa creencia ('el polvo blanco es azúcar') = resultado negativo involuntario (la compañera de Grace muere envenenada).

Hay además otros dos escenarios que operarían como "control": en uno Grace quiere envenenar a su colega y además tiene la creencia verdadera de que

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

el polvo blanco es efectivamente veneno, con lo cual su compañera muere. En el otro Grace no tiene intención de envenenar a su compañera y tiene además la creencia verdadera de que el polvo blanco es azúcar, su colega toma el café y nada malo ocurre.

Las evidencias experimentales en participantes sometidos a estos escenarios arrojan los siguientes datos:

- Los niños de cinco años o menores, al no tener aún desarrollada su área UTPD, y, por ende, al carecer de la capacidad de interpretar los estados mentales ajenos (evidencia obtenida mediante el test de la 'falsa creencia' [2]), tenderán a ver al agente que causó accidentalmente el envenenamiento como más culpable que aquel que quiso provocar la muerte y no lo logró (Young & Saxe, 2009a). Algo similar sucede en pacientes con Síndrome de Asperger (Moran et. al, 2011), quienes tienden a no evaluar diferencialmente los daños accidentales y los daños intencionales. Esto se debería a su baja capacidad para interpretar información sobre las intenciones "inocentes" de los agentes, lo que redunda en una sobre-consideración de los resultados como criterio para elaborar sus juicios morales.

- En estudios RMNf realizados en adultos normales (Young & Saxe, 2009a), se encontró una correlación significativa entre la actividad de la UTDP y la proporción de culpa asignada a Grace en las dos situaciones: a mayor actividad en la UTDP, mayor responsabilidad asignaban al agente cuando quería provocar daño aunque no lo lograra, y menor cuando causaba un daño como resultado de un accidente involuntario. Por el contrario, una menor actividad de la UTDP se correlacionaba con respuestas que tendían a fijarse menos en las intenciones y más en los resultados.
- Young y su equipo (2010a) consiguieron replicar "artificialmente" estos resultados, bloqueando temporalmente la actividad específica del área UTPD mediante estimulación magnética transcraneal (EMT). Encontraron que durante la interrupción de la actividad de la UTDP, los participantes fueron más indulgentes a la hora de juzgar los intentos fallidos de provocar daño, basándose en los resultados neutrales –vale decir, en que en última instancia no había ocurrido nada malo-; y más severos a la hora de juzgar eventos accidentales con resultados negativos.

Estos experimentos pretendían analizar la doble influencia de nuestra capacidad de interpretar estados mentales ajenos, por un lado, y de considerar los resultados de la acción, por el otro; en las competencias humanas para formular juicios morales, asumiendo que la responsabilidad moral asignada a un agente depende de la consideración de ambos factores. Ahora bien ¿qué sucede cuando (como en el caso de Mitch) el agente no tuvo ninguna mala intención, pero provocó accidentalmente –y, por ende, involuntariamente- un resultado catastrófico? ¿Tendemos, aun así, a evaluar moralmente al agente desafortunado en tales casos? Lo que Young et al. razonaron, a mi juicio, es que, si cuando procuramos interpretar los estados mentales ajenos lo que hacemos es "leer" tanto las intenciones como las creencias, en ausencia de la intención de provocar daño sólo podemos remitirnos a las creencias cognitivas del agente y usarlas como criterio para evaluar su comportamiento. En principio, estaremos más predispuestos a evaluar moralmente dichas creencias –y eventualmente las razones, también cognitivas, que llevan al agente a sustentarlas- cuando éstas dieron lugar a resultados negativos, pero Young et al. buscaron comprobar (al introducir la figura del agente "extra-afortunado") que tendemos a juzgar moralmente a otros individuos por la posesión de falsas creencias fundadas

[2] El test de la falsa creencia es usado para testear la capacidad de los niños de interpretar estados mentales ajenos: los niños contemplan una escena en la que Sally coloca una pelota en una canasta y deja el cuarto. A continuación entra otro personaje (Anne), saca la pelota del canasto y la coloca en una caja. Cuando vuelve Sally se les pregunta a los niños dónde cree Sally que está la pelota. Los niños menores de 4 años suelen responder que Sally cree que la pelota está en la caja, ya que son incapaces de representarse el estado mental de Sally como diferente del real estado de cosas.

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

en malas razones, aun cuando éstas, por efecto de un golpe extraordinario de buena suerte, no provoquen el resultado negativo esperable.

Mis críticas a la posición de los autores se remiten a dos aspectos fundamentales que desarrollaré por separado, si bien del primero pueden extraerse consecuencias relevantes para el segundo:

1º) Una crítica al enfoque estrictamente cognitivista en relación a los posibles criterios a los que apelarían los observadores cuando juzgan el comportamiento de los agentes. Argumentaré que los criterios en los que presuntamente se basan los evaluadores no se agotan (y quizás no se funden esencialmente) en el examen del valor de verdad y la justificación de las creencias del agente (aspectos puramente cognitivos que ignoran elementos motivacionales y valores implícitos).

2º) Aun cuando los juicios morales de los participantes dependieran exclusivamente del examen de las creencias y razones del agente, no es posible derivar automáticamente un juicio moral de la mera "lectura de mente" –ni siquiera de la consideración meta- cognitiva sobre el valor de verdad o la justificación de las creencias-, sin que medie alguna estimación relativa al carácter moral o inmoral (correcto o incorrecto) de las intenciones del agente. Incluso debe mediar también una valoración positiva o negativa de los resultados de la acción que permita al evaluador percibir ciertos desenlaces como afortunados o desafortunados. Esta observación sugiere que, si asumimos que el acto de valorar situaciones, personas, intenciones, creencias, etc. (tanto en términos morales como no morales) implica una tarea cognitiva per se, no reductible a la de interpretar las intenciones y creencias de un agente o los resultados de sus acciones; entonces cabría suponer que dicha tarea posee algún sustrato o correlato neural, y exige del reclutamiento de alguna región o regiones cerebrales.

Primera crítica: hacia una concepción pragmática superadora del enfoque centrado en la cognición.

Mitch deja a su hijo sólo en la bañera, basándose en la creencia de que "no se moverá". A diferencia de la creencia de Grace, cuyo contenido refiere a un estado de cosas del mundo ("el povillo blanco es azúcar" o "el polvillo blanco es veneno", según el caso) la creencia de Mitch es más bien una predicción probabilística sobre lo que él cree que podría ocurrir (o debería ocurrir) si realiza cierta acción. Ahora bien, cuando nuestras acciones se basan en creencias sobre sucesos futuros, cabe admitir que siempre existe un margen de incertidumbre, de modo tal que las mismas no suelen operar como certezas firmes y fuera de todo margen de duda. Estas creencias sólo podrían llegar a ser firmes en contextos altamente predecibles (por ej., si todos los días tomamos el colectivo en la misma parada, y éste hace siempre el mismo recorrido, y en nuestra ciudad no son habituales los cambios en los itinerarios del transporte público; entonces tendremos buenas razones para creer que mañana el colectivo parará en el mismo lugar y seguirá el itinerario de siempre). Sin embargo, en contextos en los que no podemos predecir tan fácilmente el resultado de nuestras acciones (y el escenario de Mitch, en el que se intenta predecir el comportamiento de un niño de dos años, claramente lo es) más que guiarnos por creencias firmes, lo que solemos hacer es un cálculo de probabilidades (ya sea deliberado y consciente, o rápido y automático) en virtud del cual evaluamos los posibles resultados de nuestras acciones, y la decisión que tomemos puede depender más de cierto balance entre el posible coste de asumir

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

un riesgo aun a sabiendas de la posibilidad de obtener un resultado negativo, y los beneficios resultantes de la asunción de dicho riesgo. Existen muchas evidencias de nuestra sensibilidad a los mecanismos de recompensa inmediatos o los reflejos condicionados (como el sonido del teléfono). Estos pueden desencadenar respuestas compulsivas y automáticas, dado que tenemos dificultades para reprimir el impulso a actuar en respuesta a estímulos internos (como el deseo de comer o beber) o externos (como señales que activan nuestros mecanismos de recompensa o nos ponen en estado de alerta) aun cuando seamos conscientes de que nuestra conducta puede implicar un riesgo o acarrear consecuencias negativas. Otra explicación plausible del comportamiento de Mitch, es que, si éste tomó la decisión de manera consciente y deliberada y no por influjo de un impulso irracional, no necesariamente se basó en la creencia firme e indubitable de que su hijo no se movería, sino quizás en una ponderación de valores según la cuál estimó como más relevante atender el teléfono (por ej., porque esperaba una llamada importante) que quedarse a cuidar a su hijo, aun conociendo la posibilidad de que ocurriera un accidente. Por otra parte, nuestras creencias sobre eventos futuros pueden estar profundamente sesgadas por factores subjetivos como deseos, anhelos, esperanzas, etc., dado que la voluntad de obtener resultados favorables a nuestros propósitos puede instarnos a adjudicar un alto grado de certeza a meras expresiones de deseos, lo que genera cierta confusión psicológica entre el valor cognitivo de nuestras creencias y su exclusivo sustento en la fe (por ej.: las creencias de que un tratamiento no convencional pueden curarnos de una enfermedad grave, o que una simple cábala puede ayudarnos a aprobar un examen o ganar la lotería, suelen sustentarse, no en evidencias firmes, sino actos de fe basados en el deseo de obtener resultados positivos). Del mismo modo, podríamos conjeturar que Mitch "deseaba" que su hijo se quedara quieto, y lo que guió su comportamiento fue la esperanza de que el niño le hiciera caso, más que la creencia firme de que tal cosa sucedería. Ahora bien, de ser cierto que nuestras acciones no siempre se fundan en creencias firmes, sino que, en contextos de incertidumbre (cuando no contamos con información suficiente que nos garantice el resultado a obtener) a menudo nos basamos en una ponderación implícita de los posibles resultados; también cabría esperar que quienes evalúan el comportamiento de otros agentes presupongan que éstos actúan basándose más en dicha ponderación de valores que en base a certezas cognitivas. Si a los participantes del experimento no se les hubiera dado información sobre las creencias y las razones de Mitch, sino sólo sobre sus actos y los resultados obtenidos ¿hubieran juzgado su comportamiento tomando en consideración únicamente el error cognitivo cometido por éste como consecuencia de sustentar una falsa creencia? ¿o hubieran apelado a otros criterios y argumentos a la hora de condenarlo moralmente? Mi conjetura es que los participantes podrían haber juzgado a Mitch, o bien por actuar automática e irreflexivamente en respuesta al sonido del teléfono; o bien por darle más importancia (ponderar como más valioso) al acto de atender el teléfono que a la vida, bienestar y seguridad de su propio hijo. Ambas posibilidades suponen, de algún modo, que Mitch era consciente del peligro que estaba haciendo correr a su hijo, vale decir, que no tenía la creencia firme de que su hijo no se movería, sino que sabía que ambos resultados (el negativo o el neutro) podían darse con cierta probabilidad. En tal sentido, el observador podría suponer que el comportamiento del agente no dependió –o al menos no exclusivamente- de su creencia,

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

sino, por ej., de la asunción implícita de cierta escala de valores en virtud de la cual ponderó como más importante un fin superfluo como el de atender el teléfono, antes que otro mucho más relevante: velar por la vida y seguridad de su propio hijo. Esto nos remite a una cuestión esencial que quiero recalcar en relación a mi propuesta: si los autores sostienen que lo que los observadores juzgan moralmente son sólo las creencias falsas y las malas razones de los agentes, nótese que dicho juicio no posee un contenido propiamente moral. Cuando evaluamos moralmente un acto que, como en el ejemplo analizado, no puede ser juzgado en base a las intenciones del agente, ya que no hubo intención de daño, lo que solemos hacer es, a mi juicio, suponer que dicho agente transgredió alguna norma moral cuya corrección asumimos implícitamente, norma asociada a la ponderación de valores que estimamos como prioritarios o muy importantes, en comparación con otros más superfluos, o menos relevantes. En otras palabras, cotejamos las normas y valores implícitos en el comportamiento del agente con nuestra propia jerarquía de valores, o con los principios que acatamos como guía de nuestros propios comportamientos. Eso sería, a mi modo de ver, formular un juicio propiamente moral. Si, por el contrario, asumimos que el agente en cuestión sustenta las mismas valoraciones que nosotros (en el ejemplo, pondera la vida de su hijo como el bien más preciado), pero apoya sus actos en creencias cognitivas falsas y carentes de un fundamento adecuado, lo que juzgamos es, en todo caso, la dotación intelectual del agente ("fue un estúpido al pensar que su hijo se quedaría quieto") y no sus cualidades morales. Cabe señalar que, en circunstancias 'normales' (cuando las cogniciones no redundan en acciones con malos resultados), las creencias cognitivas falsas no son objeto de juicio moral (por ej., no evaluamos moralmente a Juan por creer que afuera está lloviendo, cuando en realidad hay sol). Sin embargo, cabe admitir que cuando las creencias falsas guían acciones conducentes a desenlaces catastróficos (como accidentes que provocan daños a terceros), suele actuar el sesgo asociado a los resultados, que nos incita a juzgar moralmente al agente en la medida en que sus falsas creencias provocaron la catástrofe. Esto suele llevarnos en los hechos a confundir los planos y acusar moralmente a alguien por sus errores cognitivos, pero dicho juicio sólo posee relevancia moral en la medida en que nuestras evaluaciones morales prestan atención tanto a las intenciones del agente como a los resultados de la acción; y en presencia de malos resultados provocados accidentalmente tenderemos a juzgar cualquier factor causal que los haya desencadenado, aunque éstos no sean propiamente morales. A mi juicio, sólo el examen de los resultados podría llevar a que los evaluadores juzguen a Mitch por sus falsas creencias, pues los errores cognitivos no pueden ser inmorales per se, sino sólo por sus efectos.

Ahora bien, mi hipótesis es que existen factores propiamente morales que impulsarían a los observadores a juzgar a Mitch, y el carácter intrínsecamente moral de tales juicios remitiría, en última instancia, a la consideración del valor, status o importancia que dicho agente le adjudica a las "víctimas" de su comportamiento negligente. El término 'negligencia' posee una doble connotación, ligada a los dos tipos de factores causales de malos resultados que estamos considerando: en ocasiones se refiere a comportamientos que ponen en riesgo a otros por falta de "pericia" (lo que puede incluir malas actuaciones resultantes de la posesión de creencias cognitivas falsas), pero en la mayoría de los casos (de hecho ésta es la acepción del Diccionario de la Real Academia Española) se refiere al "descuido o falta de aplicación", lo que nos remite a su vez a la idea de desidia, desinterés y falta de preocupación por el otro. Dado que la esencia del comportamiento moral remite,

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

precisamente, a la atribución de cierto status o dignidad al prójimo, lo que nos lleva a considerar a los demás como agentes valiosos (como diría Kant, como fines per se), cabe suponer que será objeto de desaprobación moral todo comportamiento que redunde en perjuicios hacia terceros. Como en los daños accidentales el perjuicio no resulta de una acción intencional, de modo tal que no podemos apelar a las intenciones del agente; mi hipótesis es que los evaluadores asumirán que dicho perjuicio es producido indirectamente por la asunción de una escala de valores según la cual quien resulta "víctima" fue "devaluada" por el agente, vale decir, éste desestimó el valor de su vida, bienestar o seguridad por anteponer otros valores –a juicio del evaluador- más superfluos e irrelevantes. Mitch, por ej., cometió la imprudencia de correr un riesgo grave a sabiendas de las posibles consecuencias (pues difícilmente podría haber sustentado una creencia firme sobre el comportamiento de un niño tan pequeño) porque ponderó como más importante el valor superfluo de atender el teléfono que el valor, mucho más esencial, de cuidar de su hijo. Y esto con independencia del carácter deliberado o irreflexivo de dicha ponderación. Aun cuando asumamos lo segundo, la propia irreflexividad suele ser vista como un disvalor, ya que viola la norma según la cual "debemos pensar antes de actuar", y porque en última instancia dicha violación remite a la misma sensación de falta de dedicación e interés en el otro (asumimos que el cuidado de un niño pequeño exige cierto grado de concentración y aplicación si realmente nos interesa preservar su vida o bienestar). Esta valoración general del prójimo como un fin per se, cuya vida posee un valor en cierto modo incondicional, puede a su vez ser llevada a un plano más específico si lo que queremos es, por ej., correlacionar ciertos juicios de valor puntuales con su funcionalidad adaptativa, partiendo del supuesto de la Psicología Evolucionista según el cual nuestros cerebros fueron moldeados por la selección natural para experimentar emociones morales asociadas a valores cruciales para nuestra supervivencia y éxito reproductivo. Así, por ej., tendríamos cierta predisposición a proteger a los más débiles (como bebés y niños, que son incapaces de valerse por sí mismos, y además tienen toda una vida por delante, de modo que su muerte suele ser vista como más antinatural y dramática que la de un adulto o anciano); y, más aun, solemos ponderar la vida de nuestra propia descendencia como la más valiosa de todas (actitud explicable mediante hipótesis tales como la de la selección de parentesco ligada a la evolución del comportamiento altruista), de modo tal que cabe suponer que un observador juzgará como más aberrante que un padre sea responsable de la muerte de su propio hijo que de la de un desconocido. En otras palabras, es posible que el contenido específico de la "falta moral" en un escenario determinado (como el de Mitch y su hijo) no sea irrelevante e incida en el grado de severidad de la condena, dado que los observadores poseen valoraciones implícitas ligadas a contenidos concretos, las cuales pueden ser explicadas apelando a hipótesis evolucionistas, eventualmente a otros marcos teóricos.

Sin embargo, cabe admitir (quizás a favor de la tesis de Young et al.) que cuando mejor justificadas están las razones de un agente para creer que no causará daño (aun cuando sus creencias, por un golpe fortuito de mala suerte, sean falsas), menores razones tendremos para adjudicar culpa moral a dicho agente. Supongamos que el hijo de Mitch nunca antes había desobedecido a su padre en circunstancias idénticas a la descrita en el ejemplo (vale decir, éste había dejado a su hijo solo en la bañera en muchas ocasiones anteriores, y el niño nunca se había movido), o que en el lugar de trabajo de Grace el azúcar se guardaba siempre en el mismo sitio, el polvillo tenía toda la apariencia de ser azúcar y Grace no tenía la menor razón para

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

sospechar que no lo era. Sería difícil que en escenarios tan altamente predecibles (sobre todo este último), el evaluador adjudique culpa moral al agente, si éste no tenía la menor razón para creer que algo saldría mal, dadas las evidencias disponibles. Esto parece sugerir que, efectivamente, el valor de verdad de las creencias y su grado de justificación juegan un rol esencial en el juicio moral de los potenciales observadores, mientras que creencias falsas y mal justificadas desencadenarían condenas morales. Sin embargo, ésta es solo la "punta del iceberg", puesto que cuando una persona actúa en base a creencias firmes y basadas en evidencias no podemos achacarle una actitud de desidia o falta de interés (en el caso de Grace, por no corroborar si lo que había en el tarro era efectivamente azúcar; en el de Mitch, por dudar del comportamiento de su hijo cuando éste siempre le había obedecido en el pasado), mientras que quienes actúan apresuradamente en contextos de riesgo sin evidencias suficientes que permitan avalar sus predicciones (por ej., si Grace trabajara en una fábrica de productos químicos y hubiera buscado azúcar en un sector donde suelen guardarse sustancias tóxicas, hubiera tomado un tarro desconocido con un polvillo parecido al azúcar y endulzado rápidamente el café sin antes oler el producto o cerciorarse de que era azúcar) entonces serán juzgados, no tanto por sustentar una falsa creencia, sino por actuar irreflexiva y apresuradamente, lo que remite a su vez a la falta de preocupación y responsabilidad por el bienestar del prójimo. Young et al. también advirtieron la importancia de las razones del agente como uno de los aspectos más relevantes a la hora de evaluarlo moralmente, pero mi hipótesis es que la consideración de cuán justificadas son las creencias del agente no se basa en criterios puramente cognitivos; pues detrás de las malas razones los observadores podrían estar "leyendo" una falta de interés del agente por cerciorarse de que sus creencias son verdaderas, a sabiendas de que una falsa creencia puede desencadenar resultados que pongan en peligro a otros. Estos factores son los que, a mi juicio, determinan si un hecho desafortunado es caratulable (incluso en términos legales) como un caso de exclusiva "mala suerte" (lo que exime al agente de toda culpa), o si puede adjudicarse alguna responsabilidad al agente.

Segunda crítica: ¿es la interpretación de los estados mentales del agente una operación cognitiva necesaria y suficiente para la derivación del juicio moral?

Supongamos que Young et. al están en lo cierto, y lo único que los evaluadores tienen en cuenta a la hora de juzgar moralmente a Mitch son sus falsas creencias y malas razones. Tal como ellos mismos lo advirtieron, no basta con que los participantes examinen el contenido de los estados mentales (creencias y razones) en el agente, necesitan además adjudicar a dicho contenido propiedades como las de ser falso y estar débil o inadecuadamente justificado. Tales propiedades serían, por así decirlo, "meta-cognitivas", ya que no aluden directamente al contenido de las creencias del agente evaluado, sino que implican una suerte de evaluación externa de las creencias. Dado que cabe suponer que dicha operación debería tener lugar necesariamente para que las creencias sean sometidas a juicio, los autores suponen, en base a las evidencias obtenidas por RMNf, en las que no pudo detectarse una

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

activación diferencial en las regiones de interés (UTPD y UTPI) cuando los sujetos emitían juicios sobre verdad y justificación, que, o bien dichas funciones "meta-cognitivas" no son reclutadas en la UTPD o la UTPI, o bien los aparatos de medición no son lo suficientemente sensibles como para captar una posible activación diferencial ligada a tales funciones en dichas regiones de interés (ROI). Ahora bien, lo que vengo sosteniendo es que una evaluación "externa" puramente epistémica de la verdad y justificación de las creencias del agente, si bien es un paso obviamente necesario, no es requisito suficiente para la derivación de un juicio moral, pues, aun aceptando la hipótesis de que las personas condenan moralmente a otros sólo por sustentar falsas creencias basadas en malas razones (cuando éstas desencadenan daños accidentales); debería haber una operación cognitiva intermedia que lleve al evaluador a valorar negativamente, o desaprobar moralmente al agente por sustentar una falsa e injustificada creencia. He venido sosteniendo la hipótesis de que cualquier juicio moral presupone ciertas competencias del evaluador no sólo para leer las intenciones y creencias del agente, sino también para inferir los valores y normas implícitas que guían su comportamiento, a fin de ponderar en qué medida dichos valores se ajustan o se corresponden con los del evaluador, de modo tal que, cuanto más alejados estén los presuntos valores del agente de los que el observador asume como correctos o deseables, más severamente será juzgado este último. He sostenido también que, aun cuando fuera cierto que juzgamos a los agentes que cometen errores accidentales por razones "epistémicas" (falsas creencias y malas razones), este juicio puede presuponer la percepción de un "disvalor" implícito en el comportamiento del agente: por ej., podemos llegar a pensar que éste actuó precipitadamente (sin evaluar las posibles consecuencias, sin contar con evidencias suficientes para sustentar sus creencias, etc.). Estos disvalores, si bien hasta cierto punto pertenecen al terreno epistémico, cuando conducen en la práctica a malos resultados (provocan daños a terceros) suelen tener implicancias morales; no sólo porque también juzgamos las acciones por sus resultados, sino también porque podemos llegar a juzgar la negligencia como un valor negativo per se, al asociarla al desinterés y despreocupación por la vida, integridad y bienestar de los demás, disvalores que nos remiten a un terreno estrictamente moral.

Pero, aun cuando mis hipótesis sean erróneas, y las personas sólo juzguen moralmente (en los escenarios analizados) el status cognitivo de las creencias y razonamientos de otros; de todos modos parece imprescindible que el acto de evaluar moralmente suponga la existencia de una tarea cognitiva extra no reductible al mero examen de estados mentales. Aun cuando no le adjudiquemos al agente "valores", sino tan sólo creencias e intenciones (lo cual, a mi juicio, es claramente insuficiente), los observadores mismos debemos "valorar" el comportamiento y los estados mentales del agente con arreglo a algún parámetro que nos permita discriminar lo correcto de lo incorrecto, lo adecuado de lo inadecuado, lo deseable, lo preferible, etc. Si asumimos la existencia de esta tarea cognitiva, y que la misma no es reductible a otras operaciones mentales (como la lectura de mentes), un segundo paso sería intentar localizar el sustrato neural de la misma. En realidad asumo que se trata de un asunto muy com-

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

plejo, ya que cabe suponer que la operación misma de "valorar" (cualquier situación en general, y escenarios morales en particular) involucra muchas sub-operaciones (desde activaciones emocionales inmediatas hasta procesos cognitivo-deliberativos lentos y conscientes), lo que lleva a suponer que implica el reclutamiento de muchas regiones cerebrales (desde regiones más primitivas asociadas al sistema límbico, hasta las regiones más nóveles de la corteza cerebral). Por otra parte, al tratarse de una capacidad tan "global", resulta muy difícil elaborar diseños experimentales capaces de aislar la variable "valoración" de todas las sub-tareas implicadas necesariamente en el acto de valorar comportamientos, personas y situaciones (en particular la tarea de "lectura de mentes"). Esta tarea puede ser metodológicamente más sencilla cuando se trata, por ejemplo, de distinguir entre la reacción emocional provocada por la lectura de relatos o la contemplación de escenas moralmente perturbadoras (violaciones, asesinatos, etc.), de las demás funciones cognitivas involucradas en tales tareas (por ej., decodificación de los inputs visuales o de la información lingüística). Para identificar en cada sujeto experimental las regiones cerebrales responsables de otras tareas cognitivas implicadas pero que no corresponden al área específica de interés del investigador (como la decodificación visual o lingüística), primero se le pide que, por ejemplo, lea frases o vea imágenes emocionalmente neutras a fin de que la RMNf detecte la activación de esas regiones específicas, para luego; en una segunda instancia en la que se le muestran frases o imágenes ligadas a transgresiones morales con una fuerte carga emocional, la región específica de interés (en este caso regiones asociadas al procesamiento emocional de información moralmente relevante) pueda ser distinguida de las demás regiones involucradas. Ahora bien ¿cómo disociar de esta manera la actividad cognitiva de "leer la mente" (identificar creencias e intenciones de otros sujetos) con la de "valorar" dichas intenciones (o, en el caso de daño accidental, ponderar los presuntos valores del agente, o aunque más no fuera valorar sus creencias)? ¿Se trata de tareas cognitivas disociables desde el punto de vista neuronal? Tal es, a mi juicio, uno de los desafíos conceptuales y metodológicos que debería enfrentar la Psicología Moral que examina las bases neurofisiológicas de los mecanismos psicológicos asociados a la evaluación moral: identificar la región o regiones asociadas a la función cognitiva específica consistente en adjudicar a ciertos fenómenos, situaciones o comportamientos, un determinado valor (o, más específicamente, un valor moral, en el hipotético caso de que existiera un módulo o dispositivo dominio-específico para las valoraciones de este tipo).

Conclusiones

1°) Algunas sugerencias experimentales

Los experimentos de Young et. al. fueron diseñados específicamente para detectar el grado de influencia que tienen las creencias y razones sustentadas por los agentes que causan daños no intencionales, en la evaluación moral del observador (y, por ende, en la asignación de responsabilidad y culpa moral). Esto se debe a que los investigadores partieron del supuesto de que los factores específicos a los que prestan atención los evaluadores a la hora de juzgar moralmente son los estados cognitivos del agente (valor de verdad y justificación de sus creencias acerca de los posibles efectos de sus acciones). Como ya lo señalamos anteriormente, a mi juicio dichos supuestos se fundan en la hipótesis según la cual el juicio moral se

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

reduciría (o emanaría automáticamente de) la interpretación de los estados mentales del agente; y si partimos del supuesto de que tales estados se reducen a su vez a la posesión de intenciones y creencias sobre cómo llevarlas a cabo, en los casos de daños accidentales, donde no es posible atribuir intencionalidad al responsable, sólo restaría evaluar sus creencias. A este punto de vista he opuesto básicamente dos objeciones: por un lado, los participantes podrían estar prestando más atención a los valores que infieren del comportamiento del agente (y que lo llevarían a ponderar como más relevantes ciertas acciones que otras) que al contenido cognitivo de sus creencias (las cuales, como he intentado demostrar, no necesariamente son tan firmes e indubitables en contextos inciertos y relativamente impredecibles como el descrito en el escenario moral planteado). Dicha percepción de los valores implícitos en el comportamiento del agente se efectuaría tomando como punto de referencia la propia escala de valores del observador, la cual, por otra parte, es probable que goce de cierto grado de consenso social; ya sea porque los seres humanos poseemos sesgos morales innatos en respuesta a una serie de desafíos ambientales que hemos debido sortear a lo largo de nuestra historia evolutiva; o bien porque se trata de valores circulantes en el discurso social propio de un determinado contexto cultural. Así, por ej., el valor casi indiscutible que tiene en nuestra sociedad (y probablemente en cualquier cultura humana) el cuidado y protección de la prole, es tan altamente consensuado que probablemente haya llevado a los investigadores, al diseñar el experimento, a dar por sentado que un padre que deja a su hijo de dos años solo en la bañera sólo podría basar su comportamiento en la creencia firme de que el niño no sufrirá daños. Sin embargo, la realidad muestra que, pese a la alta valoración social ligada al cuidado de los hijos, no todos los padres se comportan en los hechos conforme a esos valores, pues es posible encontrar en la vida cotidiana una significativa cantidad de ejemplos de negligencia paterna ligados a la desidia, irresponsabilidad, desinterés y atención prioritaria a otros fines que desde fuera no dudaríamos en juzgar como "más superfluos". Esto sugiere además la hipótesis de que la lectura que hacemos de las intenciones o motivaciones de otros agentes depende también de nuestras experiencias previas (por ej., de la frecuencia con que solemos detectar en la vida real transgresiones de normas, o conductas que no se ajustan a valores socialmente vigentes). También, dado que la capacidad de lectura de mentes depende a su vez de nuestra capacidad de empatizar con los pensamientos y sentimientos de los demás, en la medida en que previamente percibimos las mismas inclinaciones, impulsos y emociones en nosotros mismos; el examen instrospectivo de nuestros propios deseos, que a menudo surgen, aunque no nos atrevamos a reconocerlo, de liberarnos de las exigencias del cuidado parental para ocuparnos de nosotros mismos, nos permite interpretar las mismas motivaciones en otros sujetos bajo circunstancias similares, y entender así que el comportamiento irresponsable de los padres es plausible bajo ciertas condiciones.

Mi propuesta metodológica para testear mediante un experimento conductual si los criterios que esgrimen las personas para evaluar este tipo de acciones se reducen o no al examen de los estados cognitivos del agente, o involucran otros aspectos, es muy simple: utilizaría el mismo escenario del padre que deja a su hijo en la bañera, describiendo la situación y los dos posibles resultados (el negativo y el neutro) pero

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

omitiendo hacer referencia a las creencias (verdaderas o falsas) y las razones (buenas, malas o indefinidas) del agente, de modo tal de no direccionar a los participantes para que se remitan exclusivamente al examen de los juicios cognitivos del agente. A continuación les pediría que juzguen el comportamiento de Mitch y justifiquen su evaluación moral, de modo tal de poder recabar, en una suerte de estudio preliminar, los criterios que efectivamente suelen adoptar las personas a la hora de juzgar acciones no intencionales que pueden dar lugar a malos resultados. Mi predicción es que las personas harán referencia, de un modo directo o indirecto, a ciertos defectos morales del agente (descuido, despreocupación, negligencia, falta de interés en el prójimo) ligados en última instancia a la desvalorización o subestimación del valor de la víctima. También podrían apelar a defectos como la imprudencia, precipitación, irreflexibilidad, etc., que remiten no tanto a una ponderación consciente de valores, sino más bien a un comportamiento basado en impulsos o reacciones automáticas e irreflexivas. Por último, admito que también es posible que se refieran a la "estupidez" del agente, vale decir, a cierto déficit intelectual que lo llevó a creer ingenuamente que nada pasaría (de hecho, en casos similares al del escenario planteado, a la hora de juzgar al agente se suele apelar espontáneamente al mote de "estúpido" -en referencia al error cognitivo cometido- más que a calificativos que remiten a aspectos estrictamente morales –como "desaprensivo"-). Pero, aun en el caso de enfatizar el aspecto cognitivo como el verdadero factor causal del mal resultado, también es posible, como ya ha sido señalado anteriormente, que el error del padre sea implícitamente interpretado como un signo de apresuramiento que lo habría llevado a actuar irreflexivamente sin las suficientes evidencias a favor de su creencia, o sin medir las posibles consecuencias, lo que nuevamente nos remite a un juicio moral ligado a cierta falta de interés y responsabilidad por el otro. Otra fuente de información que podría proporcionar interesante evidencia preliminar son los comentarios espontáneos de la gente sobre sucesos de este tipo, que suelen circular asiduamente en los medios periodísticos de difusión masiva a través de redes sociales como Twitter y Facebook, noticias periodísticas televisivas o publicadas on-line, etc. Un análisis del contenido de dichos comentarios (que en muchos casos contienen condenas morales muy severas) podría proporcionar pistas en relación a los criterios que suelen considerar las personas a la hora de juzgar moralmente acciones involuntarias con consecuencias negativas.

En cuanto al otro escenario paradigmático utilizado por los investigadores (el de Grace y su colega) mi sugerencia iría, de cierto modo, en sentido contrario. La diferencia principal entre ambos escenarios es que, mientras las creencias de Mitch consisten más bien en predicciones sobre sucesos futuros acerca de los que existe cierto margen de incertidumbre (y al remitirse a una situación a futuro, vale decir, que aun no sucedió, se trata de creencias en principio incomprobables); la creencia de Grace sobre la naturaleza del polvillo que tiene ante sus ojos es en principio comprobable (bastaría con fijarse bien en su textura, color, olor, o, a lo sumo, hacer alguna prueba más sofisticada para comprobar de qué sustancia se trata).

Los investigadores apelaron al ejemplo de una Grace extra-afortunada quien, pese a sustentar la creencia errónea de que el polvo blanco encontrado en una alacena era azúcar (cuando en realidad, era veneno),

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

endulzar con dicho polvo el café de su compañera y servírselo, tuvo la buena suerte de que ésta última lo dejara a un lado y se olvidara de tomarlo, con lo cual el evento desafortunado (muerte por envenenamiento) no ocurrió. Young et. al. conjeturaron que la extra-afortunada Grace sería juzgada por su falsa creencia aun en ausencia del resultado desafortunado, y efectivamente eso fue lo que ocurrió en la situación experimental. Tal como lo he venido sugiriendo anteriormente, mi predicción para este escenario es que el grado de severidad o indulgencia en el juicio de los participantes dependerá, en consonancia con lo que sugieren Young et. al., del grado de justificación de la creencia de Grace: a medida que aumentan las evidencias a favor de los motivos de Grace para sustentar una creencia firme en relación a la naturaleza del polvillo blanco (en el límite, en ausencia total de razones válidas que puedan haberla llevado a sospechar que éste no era azúcar -porque todos los días extrae el azúcar del mismo recipiente, porque no hay motivos para creer que alguien puede haber tenido la intención de modificar su contenido, etc.-) la condena moral se reduciría a cero, y los participantes deberían juzgar el hecho como puramente accidental, eximiendo a Grace de toda culpa. Por el contrario, la adjudicación de culpa moral debería ir en aumento cuanto más débil sea la creencia de Grace de que "el polvillo es azúcar" (como ya lo sugerimos, si, por ej., Grace trabajara en una fábrica de productos tóxicos, si hubiera buscado azúcar en un sector donde suelen guardarse otras sustancias químicas, si hubiera tomado un recipiente de origen desconocido y no se hubiera cerciorado por algún medio de la naturaleza de su contenido, etc.). En tal sentido, mi sugerencia para el diseño experimental consistiría en presentar a los participantes distintas variantes del escenario moral en las que lo que varíe sea el grado de certeza atribuible a la creencia de Grace en función de las razones y evidencias disponibles (algo similar a las razones buenas o malas del experimento de Mitch). En cada uno de esos escenarios, les pediría que respondan, utilizando una escala similar a la de los investigadores, cuán moralmente culpable es Grace por haberle servido café envenenado a su colega (aun en ausencia del resultado negativo), y que justifiquen su respuesta. Mi predicción es que en el caso de las creencias débilmente justificadas los participantes harán referencia a "disvalores" tales como la negligencia, el descuido, la falta de atención, etc., a los cuales les he asignado en este trabajo un contenido propiamente moral (a diferencia de las creencias cognitivas per se, que no serían ni morales ni inmorales). Si, en cambio, el agente que cometió el error amparándose en una creencia lo suficientemente bien justificada no es moralmente juzgado, esto aportaría alguna razón de peso para pensar en que en tales casos no es posible adjudicar defectos como descuido, precipitación o negligencia a quienes actúan conforme a creencias bien fundadas. En otras palabras, mi conjetura es que los participantes evaluarán el riesgo potencial de la acción de Grace en función del contexto, y, por ende, la tendencia del agente a asumir riesgos que pueden poner en peligro a terceros, lo que indirectamente remite a la ponderación implícita del valor de la vida y la seguridad de los demás. Reconozco que cabe la posibilidad de que tales criterios morales, aun cuando eventualmente formen parte de las inferencias implícitas de los participantes, no sean explicitados por ellos (en cuyo caso no podríamos estar seguros de que efectivamente forman parte de sus razonamientos tácitos), pero creo que ésta puede ser una posible vía para

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

explorar de manera un poco más sutil de qué modo ciertos factores ligados al contexto en que se produce la toma de decisión pueden (o no) afectar las evaluaciones morales.

En esta revisión también he sugerido la importancia de idear experimentos mediante el uso de técnicas de RMNf a los fines de detectar si se producen activaciones diferenciales en alguna región o regiones cerebrales cuando los participantes:

1º) Procuran identificar o interpretar los valores subyacentes al comportamiento del agente, bajo el supuesto teórico de que los observadores no sólo basan su evaluación moral en el examen de las creencias e intenciones de los agentes, sino que infieren implícitamente aspectos ligados a su axiología: lo que les importa más o menos, les resulta más o menos preferible, más o menos deseable o estimable, etc.

2º) Realizan sus propios juicios morales, lo que implica que ellos mismos estarían apelando a su propia escala de valores para ponderar el carácter correcto o incorrecto del comportamiento del agente. Hemos conjeturado que la operación de "valorar" debería tener un status cognitivo per se, no reductible a la mera interpretación de creencias e intenciones.

En lo atinente al primer punto, la utilización de escenarios asociados a la "suerte moral" puede ser un camino promisorio para identificar qué es lo que los sujetos experimentales evalúan moralmente cuando no existe la intención de provocar daño. La conjetura que sostienen Young et. al. y procuraron contrastar experimentalmente es que, en tales casos, lo único que queda por evaluar son las creencias y razones del agente, vale decir, aspectos puramente cognitivos. En tal sentido, creo que los investigadores cometen una suerte de "falacia cognitivista" (muy ligada al enfoque de la Psicología Cognitiva, dominante en los actuales abordajes en Psicología Moral experimental). La falacia cognitivista en el terreno filosófico está más asociada a la ética normativa: consiste en la suposición de que es posible derivar juicios valorativos y normativos a partir de juicios de conocimiento. El enfoque de los autores es, en cambio, descriptivo: lo que están afirmando es que los evaluadores derivan *de hecho* consecuencias morales (juicios de valor) del examen del contenido cognitivo de las creencias del agente (en particular, de su valor de verdad y su grado de justificación, aspectos meramente epistémicos). Lo que yo conjeturo, en cambio, es que estos aspectos no son los únicos tenidos en cuenta por los evaluadores a la hora de juzgar moralmente (más aun, no son los más relevantes), y que, en todo caso, tras los juicios de los participantes sobre cuán justificadas son las creencias del agente subyace la atribución a estos últimos de cierta escala de valores que el participante percibe como "trastocada" (al ponderar como más importantes valores "superfluos" por encima de otros objetivamente considerados más relevantes). En otras palabras, el evaluador le asigna valores al agente. Ahora bien, ¿cómo comprobar si la activación de ciertas regiones de interés cuando los sujetos emiten sus juicios morales corresponde a la asignación de creencias y razones al agente, o a la asignación de valores? La única estrategia que se me ocurre es crear escenarios que expliciten los valores y preferencias del agente (por ejemplo: "Juanita, una madre adolescente, desea con todas sus fuerzas ir a bailar, de modo tal que decidió dejar a su hijo de 9 años solo al cuidado de sus dos hermanitos de un año y medio y 4 meses

[3] Ejemplo inspirado en un caso de la vida real: una joven argentina de 21 años fue a bailar una noche dejando solos a sus cuatro hijos de 6, 4, un año y medio y cuatro meses (supuestamente todos al cuidado del más grande). El bebé de cuatro meses falleció ahogándose con la mamadera.

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

para poder ir a la disco con sus amigas" [3]), y en otra prueba describir el mismo escenario, pero explicitando las creencias y razones del agente ("Juanita creía que su hijo de 6 años podía cuidar a los demás y que no se producirían incidentes, porque ya antes los había dejado solos"). Si bien esta propuesta puede generar complicaciones (es claro que ambas descripciones instarían a la condena moral de Juanita) al menos quizás podría comprobarse si se produce una activación diferencial cuando las personas examinan las valoraciones y preferencias del agente, en comparación con el examen de sus creencias y razones.

En cuanto al segundo aspecto, creo que es bastante más complejo aun. Si lo que queremos probar es que el acto de evaluar moralmente implica un plus no reductible a la mera lectura de mentes, lo que podría hacerse es, de manera preliminar, someter a los sujetos experimentales a un primer experimento en el que se le proporcionan frases que describen acciones moralmente neutras (por ej.: "Pedro quiere ir de paseo al campo pero sabe que allí hay muchos mosquitos, por lo tanto, Pedro compra antes un repelente de insectos"). Se supone que el examen del comportamiento de Pedro remitirá al sujeto experimental a la lectura de sus intenciones ("ir al campo") y sus creencias ("en el campo hay muchos mosquitos"), por lo tanto, deberían activarse las regiones asociadas a la lectura de mente (UTPD). En una segunda instancia, se le proporcionan a los mismos sujetos frases que describen acciones moralmente evaluables (por ej., "Pedro quiere matar al hijo del vecino porque le rompió un vidrio de la ventana de un pelotazo. Pedro sabe a qué hora el hijo del vecino se queda sólo en la casa, y que desde la ventana de su cuarto puede disparar un arma y acertarle al niño justo en la cabeza. Pedro espera que se vaya su vecino, toma su pistola calibre 22 y desde la ventana de su cuarto le dispara al niño"). Si en este ejemplo los participantes, además de examinar las intenciones y las creencias de Pedro realizan otra tarea cognitiva extra (la de valorar moralmente tales intenciones), cabe esperar que se activen otras regiones del cerebro, o acaso que haya una mayor activación en la misma región (UTDP), en fin, cabría esperar en términos generales encontrar una activación diferencial con respecto al primer experimento. Esto estaría dando la pauta de que el acto de valorar moralmente requiere de la lectura de mentes como condición necesaria pero no suficiente (vale decir, implica la realización de tareas cognitivas subyacentes, tales como la capacidad para interpretar los estados mentales de otros agentes, pero no es reductible a dicha tarea). En otras palabras, evaluar moralmente no es sólo leer los estados mentales del agente, sino que implica además la operación metacognitiva de valorar tales estados.

2°) Algunas implicancias en el terreno de la praxis

Los estudios experimentales en Psicología Moral (que apuntan, en términos generales, a desentrañar los mecanismos psicológicos -y sus correlatos neuronales- intervinientes en la producción de juicios y evaluaciones morales y en el razonamiento moral en general), poseen, como en el caso específico de la "suerte moral", implicancias prácticas en diversos

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

frentes. Así, por ejemplo, un examen científico de cómo las personas suelen evaluar acciones no intencionales con resultados catastróficos como producto de la "mala suerte", podría afectar nuestra percepción de la corrección o incorrección de los fallos judiciales en los que se imputa a algún sujeto, grupo, empresa, etc., por las consecuencias negativas involuntarias de sus acciones. Las propias legislaciones en materia penal también incluyen criterios sobre cómo evaluar este tipo de hechos. Una de las tantas ramas actuales de la neurociencia es el Neuroderecho, disciplina que se ocupa de reflexionar sobre los posibles usos jurídicos de los avances neurocientíficos. Tales avances pueden ser utilizados como herramientas (tratamientos "correctivos" neurofarmacológicos, pruebas basadas en imágenes cerebrales capaces de aportar datos sobre la condición psíquica del imputado, o sobre la veracidad de su declaración), o bien como evidencias teóricas que presuntamente contribuirían a la modificación de ciertas concepciones jurídicas, ejerciendo una influencia decisiva sobre el sistema judicial y el aparato legal en general, o bien sobre los criterios de los jueces a la hora de prescribir una pena, asignar culpabilidad o declarar la imputabilidad o inimputabilidad de un acusado.

Ahora bien, ¿en qué medida nuestro conocimiento del modo como de hecho producimos juicios morales puede influir sobre lo que deberíamos hacer en materia jurídica y legal, o, en términos más generales, sobre el modo como deberíamos juzgar a las personas y sus actos? Nos enfrentamos aquí al clásico problema de la "guillotina de Hume", según el cual no es posible derivar ninguna consecuencia normativa de nuestro conocimiento de lo que las cosas son (Hume, 1751), en este caso, del modo como nuestro cerebro procesa información moralmente relevante, puesto que, tal como lo señala Moore (1903), las propiedades naturales no entrañan ninguna condición intrínseca de 'buenas' o 'malas'. Para atribuirles tal condición debemos inevitablemente asignarles un valor "desde fuera", para lo cual nos valemos de una suerte de reflexión "de segundo orden", que no es otra cosa que el ejercicio de la crítica racional, fundamento de cualquier Ética normativa. Ahora bien, lo más enigmático del asunto es justamente ese desdoblamiento entre "lo que somos", o el modo como evaluamos moralmente de manera "espontánea" o "natural" (si es que cabe tal cosa, puesto que el razonamiento moral es a su vez producto de nuestras experiencias e historia de vida, atravesada por múltiples influencias ambientales, familiares y socio-culturales); y nuestra capacidad de tornarnos conscientes de nuestras propias disposiciones y sesgos psicológicos y examinarlos críticamente, desde una suerte de perspectiva "externa", de 'tercera persona'. Dicha capacidad de razonamiento crítico de "segundo orden" debería poder ser también objeto de una explicación naturalista (después de todo, se trata de una capacidad que forzosamente debería estar también en nuestro cerebro), pero esto a su vez debería permitir la emergencia de una autorreflexión de 'tercer orden', generada también por nuestro cerebro, y así ad infinitum.

Dejando de lado semejante problema filosófico, si nos propusiéramos de todos modos examinar las implicancias normativas de los descubrimientos en Psicología Moral, caben dos posibilidades opuestas, lo que demuestra una vez más la imposibilidad de derivar juicios de valor unívocos y definitivos como consecuencia

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

de las evidencias experimentales (las cuales, a su vez, son falibles y siempre están sujetas a futuros análisis y revisiones). Dichas posibilidades son compatibles con dos fuertes tradiciones filosóficas: el emotivismo humeano y el racionalismo kantiano:

- Por un lado, podríamos argüir, siguiendo una versión evolucionista aggiornada de la tradición naturalista inaugurada por Hume, que nuestros juicios morales son "correctos" y "adecuados" en la medida en que resultan de una arquitectura cerebral que evolucionó para responder eficazmente a ciertos desafíos adaptativos del ambiente social, de modo tal que las emociones implicadas en nuestras evaluaciones morales (en el caso analizado, por ej., la emoción aversiva que despierta un mal resultado, incitándonos a culpar moralmente al responsable) serían herramientas adecuadas para estimular juicios moralmente "válidos" (entendiendo por validez, en este caso, alguna suerte de eficacia adaptativa para el colectivo social y, en consecuencia, para cada individuo que lo integra). Desde esta perspectiva, lo que somos de hecho (nuestras competencias morales y los juicios de valor que promueven) es considerado a priori como 'bueno', bajo el supuesto de que tales juicios son subsidiarios de capacidades cognitivas y emocionales que "están allí" porque se fijaron en nuestro repertorio genético o en nuestra arquitectura cerebral en razón de su valor para resolver algún tipo de problema adaptativo. En tal sentido, sería correcto derivar el deber-ser de lo que las cosas son.

- En la vereda opuesta, podríamos considerar que las evaluaciones morales espontáneas de las personas (por ej., la tendencia, confirmada hasta cierto punto por Young et. al., a condenar moralmente a un agente sólo, o fundamentalmente, cuando las consecuencias de sus actos fueron negativas), están sesgadas por factores emotivos que distorsionan el juicio objetivo e imparcial de los hechos. Mientras para la posición anterior tal tendencia, aunque aparentemente irracional, tendría su razón de ser o justificación en el hecho de que los malos resultados pueden afectar crucialmente la supervivencia, integridad o bienestar de personas, objetos, espacios físicos, etc., de modo tal que sería razonable experimentar una fuerte emoción de rechazo ante tales "errores humanos" que condicionaría profundamente nuestros juicios explícitos; desde esta perspectiva dichos sesgos serían contrarios a los dictámenes de la razón aun admitiendo su posible valor adaptativo. En tal sentido, la utilidad práctica de las evidencias experimentales sobre nuestras competencias para el juicio moral residiría en ayudarnos a detectar, someter a juicio crítico y eventualmente eliminar los sesgos, falacias o errores cognitivos implicados en nuestras evaluaciones morales. El deber-ser iría en este caso a contrapelo de lo que las cosas son, y se traduciría en el clásico mandato kantiano de obrar conforme a la razón y contra toda inclinación (Kant, 1785).

Un análisis profundo de las implicancias de estas tesis opuestas excede los límites de este trabajo. Sólo diré brevemente que ambas, paradójicamente, expresan posiciones compatibles con dos tendencias propias del razonamiento humano que tienen sus respectivas expresiones en regiones cerebrales distintas: el utilitarismo, que nos insta a juzgar conforme a los resultados (si algo es útil y aporta mayores beneficios que perjuicios, es moralmente aceptable); y el deontologismo (lo único válido en términos morales son las buenas intenciones del agente, y si algo es propiamente moral,

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

lo es porque posee un valor intrínseco e independiente de sus consecuencias). Es curioso este fenómeno de la recursividad, el extraño bucle del que hablaba Hofstadter (2007), en virtud del cual encontramos en el cerebro las propias herramientas que utilizamos para abordarlo, lo que entraña una circularidad insalvable. En este trabajo he intentado argumentar, desde una perspectiva naturalista, que el enfoque adoptado por Young et. al. (impregnado del prejuicio cognitivista dominante en las ciencias cognitivas) fracasa al eludir el aspecto crucial de todo juicio moral: la atribución de valores implícitos en el comportamiento del agente y la ponderación de tales valores por parte del evaluador. En otras palabras, intenté mostrar que el acto de valorar (moralmente o no) todo lo que nos rodea: situaciones, personas, comportamientos, hechos, objetos, etc., es una parte constitutiva del psiquismo humano irreductible a cualquier descripción de lo hechos, de modo tal que la información meramente cognitiva nada nos dice sobre cómo estimar aquello que tenemos delante. Esta conclusión es, por ende, aplicable al tema que nos ocupa ahora: no hay modo de derivar unívocamente consecuencias normativas de las evidencias empíricas sobre el funcionamiento de nuestro cerebro, y si las derivamos, tales consecuencias sólo pueden desprenderse de los datos experimentales en la medida en que asumimos una posición filosófica a-priori, ella misma valorativa.

Tales criterios axiológicos no sólo se ponen en evidencia a la hora de derivar consecuencias normativas de los datos empíricos, también están presentes en las asunciones implícitas de los investigadores al momento de diseñar el experimento. Aun cuando el objetivo de estos estudios es de índole descriptivo (vale decir, de lo que se trata es de formular conjeturas sobre cómo la gente razonaría de hecho frente a ciertos escenarios morales) el sólo hecho de intentar someter a prueba algún tipo de "falacia" (ya sea postulada a priori o identificada a posteriori del experimento), supone que los investigadores deben asumir de antemano ciertos criterios en relación a lo que debería ser una evaluación moral adecuada, fundada sobre bases racionales (una suerte de modelo implícito de optimización de los juicios morales, a la manera de los modelos de optimización usados en campos como la Biología Evolucionista, la Economía, etc.) y tomarlos como parámetro a fin de comprobar si los juicios espontáneos de los participantes se ajustan o no a dichos supuestos.

Received: 27/11/2014
Accepted: 30/03/2015

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

Referencias

Gray, K. Young, L., Waytz, A. 2012. Mind Perception Is the Essence of Morality, Psychological Inquiry, 23:101–124.
Haidt, J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. Psychological Review 108: 814–834.
Hofsdtadter, D. 2007. Gödel, Escher, Bach: un eterno y grácil bucle. Barcelona: Tusquets Ed.
Hume, D. 1751. Investigación sobre los principios de la moral. Madrid: Alianza, 2006.
Kant, I. 1785. Fundamentación de la metafísica de las costumbres. Madrid: Tecnos, 2006.
Knobe, J. 2005. Theory of mind and moral cognition: Exploring the connections. Trends in Cognitive Sciences 9: 357–359.
Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., & Gabrieli, J. D. 2011. Impaired theory of mind for moral judgment in high-functioning autism. Proceeding on the national Academy of Sciencies, 108, 2688-2692.
Moore, G. E.1903, Principia Ethica. Barcelona: Crítica, 2002.
Nagel, T. 1979. "Moral luck". Mortal questions, 24–38. Cambridge: Cambridge University Press.
Richards, N. 1986. Luck and desert. Mind 65:198-209.
Rosebury, B. 1995. Moral responsibility and moral luck. Philosophical Review 104:499-524.
Royzman, E., and R. Kumar. 2004. Is consequential luck morally inconsequential? Empirical psychology and the reassessment of moral luck. Ratio 17: 329–344.
Williams, B. 1982. "Moral luck." Moral luck, 20–39. Cambridge: Cambridge University Press.
Young, L., Cushman, F., Hauser, M., Saxe, R. 2007. The neural basis of the interaction between theory of mind and moral judgment, Proceedings of the National Academy of Sciences 104(20): 8235–8240.

Volumen 9. Número 1. Abril 2015. DOI: 10.7714/cnps/9.1.202

Young, L., and R. Saxe. 2008. The neural basis of belief encoding and integration in moral judgment. NeuroImage 40: 1912–1920.

Young, L., and R. Saxe. 2009a. Innocent intentions: A correlation between forgiveness for accidental harm and neural activity. Neuropsychologia 47: 2065–2071.

Young, L., Saxe, R.. 2009b. An fMRI investigation of spontaneous mental state inference for moral judgment. Journal of Cognitive Neuroscience 21: 1396–1405.

Young, L., Nichols, S., Saxe, R. 2010a, Investigating the Neural and Cognitive Basis of Moral Luck: It's Not What You Do but What You Know. Rev.Phil.Psych.1:333–349.

Young, L, Camprodon J. A., Hauser M., Pascual-Leone, A., Saxe, R., 2010b. Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. Proceedings of the National Academy of Sciences 107(15): 6753–6758.

Young L, Dodell-Feder D, Saxe R. 2010c. What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. Neuropsychologia 48(9):2658-64.

Young, L., Scholz, J., Saxe, R. 2011, Neural evidence for "intuitive prosecution": The use of mental state information for negative moral verdicts, Social Neuroscience 6(3):302-315.

Zavadivker, M. N. 2014a. Entre el naturalismo y la normatividad: los dilemas bioéticos bajo el scanner, Cuadernos de Neuropsicología 8 (1):20-43.

Zavadivker, M. N. 2014b. Homo eticus. Las bases biológicas del comportamiento pro-social, San Miguel de Tucumán: Edic. La Monteagudo.