

# *Hacia una neurobiología cognitiva de las virtudes morales\**

PAUL M. CHURCHLAND

*Universidad de California, San Diego*

## I. INTRODUCCIÓN

Son estos los primeros días de lo que espero que sea una tradición intelectual larga y fructífera; una tradición estimulada por la interacción sistemática y la mutua información entre la neurobiología cognitiva por un lado y la teoría moral por el otro. Más específicamente, es el sub-área tradicional que llamamos *metaética*, incluyendo a la epistemología moral y a la psicología moral, la que resultará informada de forma más dramática por los amplios desarrollos de la neurobiología cognitiva. Y es también la metaética la que ejercerá una influencia recíproca sobre la investigación neurobiológica futura —más específicamente, sobre la naturaleza de la percepción moral, la naturaleza del razonamiento práctico y social, y el desarrollo y corrupción ocasional del carácter moral.

Este último punto sobre la reciprocidad destaca un punto adicional. Lo que estamos contemplando aquí no es una toma imperialista de lo moral en manos de lo neuronal. Más bien debemos prever un florecimiento mutuo *tanto* de nuestras ideas de alto nivel en el dominio del conocimiento moral *como* de nuestras ideas de más bajo nivel en el dominio de la neurología normal y patológica. Pues cada nivel tiene mucho que enseñar al otro, como intentará mostrar este ensayo.

Tampoco tenemos que resistirnos a esta interacción entre distintas tradiciones sobre la base de que amenaza con deducir conclusiones normativas a partir de premisas puramente fácticas, pues no amenaza con tal cosa. Para verlo claramente, considérese el siguiente paralelismo. La neurobiología cognitiva está también en el proceso de arrojar mayor luz

\* Traducción de Antonio Diéguez.

sobre la filosofía de la *ciencia* –mediante el desvelamiento de las diferentes formas de representación neuronal que subyacen a la cognición científica, y de las diferentes formas de actividad neuronal que subyacen en el aprendizaje y el cambio conceptual (véase, por ejemplo, Churchland 1989, capítulos 9-11). Y sin embargo, son los científicos los que continuarán haciendo la ciencia real, según los diversos métodos mediante los cuales se realiza el progreso científico. Una teoría adecuada del cerebro no constituirá evidentemente una teoría de la Evolución Estelar o una teoría de la estructura subyacente de la Tabla Periódica. Constituirá a lo sumo una teoría de cómo generamos, expresamos y manipulamos esos valiosos logros cognitivos.

Del mismo modo, y por las mismas razones, serán los pensadores morales y políticos los que continuarán haciendo la teoría moral y política reales, según los diversos métodos mediante los cuales se realiza el progreso moral y político. Una teoría adecuada del cerebro no constituirá obviamente una teoría de la Justicia Distributiva o un Código Penal. Constituirá a lo sumo una teoría de cómo generamos, expresamos y manipulamos esos valiosos logros cognitivos.

Podría parecer que estas promesas tranquilizadoras despojan de interés al citado programa, al menos para el filósofo moral, pero veremos inmediatamente que no es así. Pues vamos a contemplar una explicación sistemática y unificada, esbozada en términos de redes neuronales, de los siguientes fenómenos: el conocimiento moral, el aprendizaje moral, la percepción moral, la ambigüedad moral, el conflicto moral, el razonamiento moral, las virtudes morales, el carácter moral, la patología moral, la corrección moral, la diversidad moral, el progreso moral, el realismo moral y la unificación moral. Este esbozo colectivo servirá al menos para perfilar el programa, e incluso en esta fase inicial proporcionará una base desde la que establecer las credenciales de una destacada tendencia en la metaética pre-neuronal: el programa conocido como «Ética de la Virtud», representado tanto por un autor antiguo (Aristóteles), como por tres autores modernos (Johnson, Flanagan y MacIntyre).

## II. LA RECONSTRUCCIÓN DE LOS FENÓMENOS COGNITIVOS MORALES EN TÉRMINOS NEUROBIOLÓGICOS COGNITIVOS

Este ensayo se cimenta en trabajos que ya tienen aproximadamente una década, trabajos concernientes a la capacidad de los recientes modelos de redes neuronales (de la actividad cerebral de micronivel) para reconstruir, en forma explicativa, los rasgos principales de la actividad cognitiva de nivel global. Esta investigación comenzó a mitad de los ochenta estudiando los problemas del reconocimiento perceptivo, la gé-

nesis de la conducta motora y otros fenómenos básicos que implicaban el aprendizaje gradual de diversas *habilidades* cognitivas por redes «neuronales» artificiales modeladas con grandes computadoras digitales (Gorman y Sejnowski, 1988; Lehky y Sejnowski, 1988; Rosenber y Sejnowski, 1990; Lockery et al., 1990; Cottrell, 1991; Elman, 1992). A partir de ahí ha adoptado un enfoque descendente, para intentar entender con detalles más fieles la estructura empírica del cerebro biológico (Churchland y Sejnowski, 1992), y también un enfoque ascendente, para entender la estructura y dinámica de fenómenos cognitivos de alto nivel, como los que se muestran, por ejemplo, en el interés humano por las diversas ciencias teóricas (Churchland, 1989).

Para los filósofos, la introducción más rápida y fácil a estas ideas generales quizás sea la explicación ampliamente ilustrada de Churchland (1995), a la cual remito al lector lego. No es mi objetivo aquí recapitular esta obra preliminar, sino construir sobre ella. No cabe duda, sin embargo, de que, incluso para el lector al que le resulten nuevas estas ideas, la información de fondo emergerá lentamente en los diversos ejemplos que siguen, así que simplemente continuaré mi exposición y confío en que la cosa irá bien.

El modelo que seguiré aquí es mi anterior intento de reconstruir la epistemología de las ciencias *naturales* en términos de redes neuronales (Churchland, 1989). Mi propio interés filosófico ha estado siempre centrado en cuestiones de epistemología y de filosofía de la ciencia, de modo que era natural, a mediados de los 80, que intentara primero aplicar el marco emergente de la neurobiología cognitiva a las cuestiones que me eran más familiares. Pero pronto se me hizo evidente que este marco emergente poseía una generalidad inesperada, y que su poder explicativo, si era auténtico, iluminaría una extensión de fenómenos cognitivos mucho mayor de la que hasta entonces había sido tratada. Así pues, propuse extender su aplicación a otras áreas cognitivas, tales como el conocimiento matemático, el conocimiento musical y el conocimiento moral. (Algunas primeras incursiones aparecen en los capítulos 6 y 10 de Churchland, 1995). Aunque solo fuera eso, estos dominios adicionales de actividad cognitiva proporcionan una serie de *pruebas* difíciles para las suposiciones y ambiciones explicativas de la teoría de redes neuronales. Por tanto, el presente trabajo pretende extraer las tesis teóricas centrales a las que nos compromete un modelo de cognición de redes neuronales en el dominio de la metaética. El lector, y especialmente el filósofo moral profesional, juzgará si el retrato general resultante es instructivo desde un punto de vista explicativo y fiel a la realidad moral.

## 1. El conocimiento moral

Hablando en términos generales, enseñar o entrenar a una red neuronal para que encarne una capacidad cognitiva específica es imponer gradualmente una *función* específica a su comportamiento de entrada-salida. La red adquiere así la capacidad para responder, de formas diversas pero sistemáticas, a una amplia variedad de entradas sensoriales potenciales. En una red simple de tres capas, alimentada hacia adelante (*feedforward*) y con conexiones sinápticas fijas (figura 1a), el comportamiento de salida en la tercera capa de neuronas está completamente determinado por la actividad de la capa de entrada sensorial. En una red *recurrente* (biológicamente más realista) (Figura 1b), el comportamiento de salida está determinado conjuntamente por la entrada sensorial y por el estado dinámico previo de toda la red. El caso de la red alimentada sólo hacia adelante depara una capacidad cognitiva que es sensible a patrones espaciales, pero es ciega a patrones temporales o al contexto temporal; la red recurrente depara una capacidad cognitiva que es sensible, y capaz de reaccionar, a contextos cognitivos cambiantes en los que las entradas sensoriales son recibidas de formas variadas. En ambos casos, la capacidad cognitiva adquirida reside realmente en la configuración específica de las múltiples *conexiones* sinápticas entre las capas neuronales, y aprender esta capacidad cognitiva es una cuestión de ajustar lentamente el tamaño o «peso» de cada conexión, de modo que lleguen a encarnar colectivamente la función de entrada-salida que se desee. Volveremos después sobre ello.

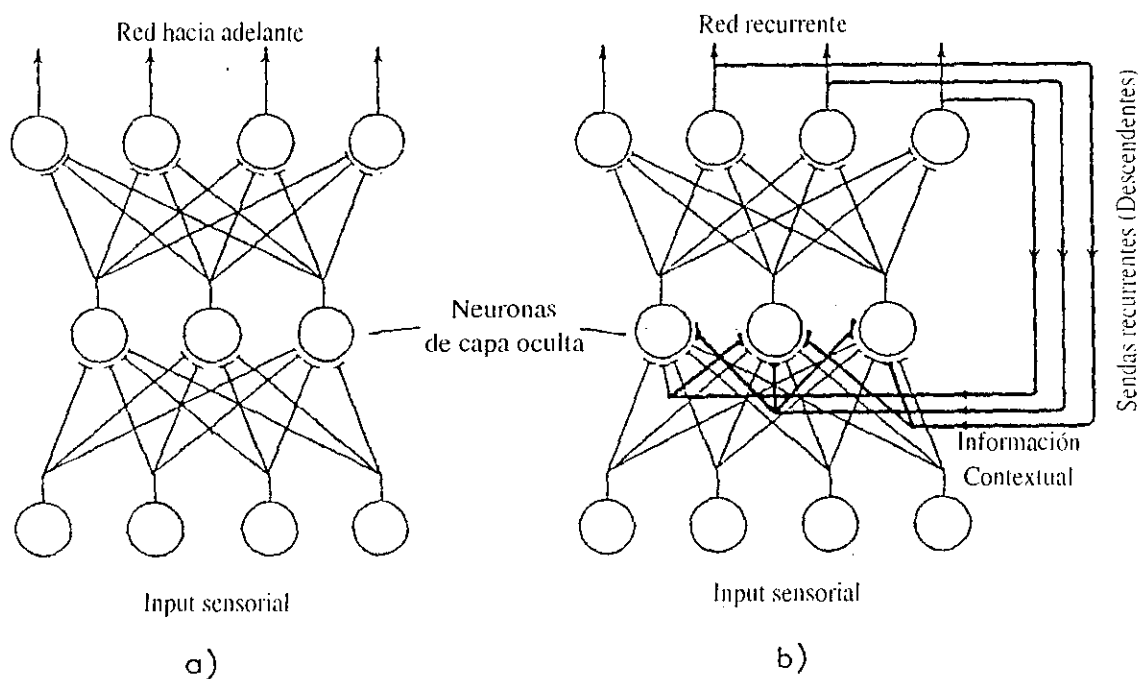


Figura 1. (a) Una red simple alimentada hacia adelante. (b) Una red recurrente simple. Para una rápida comprensión del significado funcional de estos modelos, piénsese en la capa de neuronas inferior o de entrada como si fueran neuronas sensoriales, y piénsese en la capa de neuronas superior o de salida como si fueran neuronas motoras o activadoras del músculo.

Evidentemente, una red entrenada ha adquirido una habilidad específica. Es decir, ha aprendido a responder a las diversas entradas de su capa sensorial con patrones apropiados de actividad neural a través de su capa de salida. Por consiguiente, mi primera caracterización del conocimiento moral, al igual que cualquier otro modo de conocimiento, es como un *conjunto de habilidades*. Para empezar, un adulto moralmente entendido ha adquirido claramente una sofisticada familia de habilidades *perceptivas* o de *reconocimiento*, las cuales le permiten una comprensión fluida de sus propias circunstancias sociales y morales, y de las circunstancias sociales y morales de los otros en su comunidad. E igualmente claro es que un adulto moralmente entendido ha adquirido un complejo conjunto de habilidades *conductuales y manipulativas*, las cuales posibilitan su interacción social y moral exitosa con los otros en su comunidad.

Según el modelo de cognición que estamos aquí explorando, las habilidades en cuestión están encarnadas en una vasta configuración de conexiones sinápticas con los pesos adecuados. Seguramente no sea intuitivamente obvio cómo un millar, o un millar de millones, o un billón de tales conexiones pueden constituir una habilidad cognitiva específica, pero comenzamos a captar intuitivamente cómo lo pueden hacer cuando prestamos atención al comportamiento colectivo de las neuronas de la capa a la que pertenecen estas conexiones cuidadosamente configuradas.

Considérese, por ejemplo, la segunda capa de la red alimentada hacia adelante en la Figura 1a. Esta población neuronal, como cualquier otra población neuronal discreta, representa las diversos estados del mundo con una correspondiente variedad de *patrones de activación* a través de toda la población. O sea, así como un patrón de niveles de brillo a través de los 200.000 píxeles de su familiar pantalla de TV puede representar una determinada escena bidimensional, del mismo modo puede el patrón de niveles de activación a través de una población neuronal representar aspectos específicos del mundo externo, aunque sólo rara vez la «semántica» de esta relación representacional será tan claramente «pictórica». Si la representación neuronal es, por ejemplo, auditiva, u olfativa, o gustativa, entonces la representación será obviamente distinta a una «imagen» 2-D.

Lo que es importante para nuestros propósitos es que el *espacio* abstracto de patrones representacionales *posibles*, a través de una población neuronal dada, adquiera lentamente en el curso del entrenamiento de las

sinapsis una estructura específica –una estructura que, en respuesta a una familia preferida de estímulos distintos en la capa sensorial de la red, asigne una familia de *localizaciones* abstractas espectacularmente preferenciales en este espacio. Así es como la red madura se las arregla para categorizar todas las posibles entradas, o bien como instancias toscas de una-u-otra de su familia de *categorías* prototípicas aprendidas, o bien, si esto falla, como instancias de ruido ininteligible. Antes del entrenamiento *todas* las entradas producen ruido en la segunda capa. Sin embargo, después del entrenamiento, esta segunda capa se ha sensibilizado con preferencia para un subconjunto comparativamente pequeño del amplio alcance de posibles patrones de entrada (a la mayor parte de los cuales nunca se los encuentra). Cuando ocurren estos patrones de entrada «cruciales», son asimilados en el conjunto adquirido de *categorías prototípicas* de la segunda capa.

Considérese una red artificial (Figura 2a) entrenada para discriminar caras humanas de no-caras, caras masculinas de caras femeninas, y un puñado de individuos designados entre una variedad de fotografías distintas. Como resultado de este entrenamiento, el espacio abstracto de patrones de activación *posibles* ha sido primero *dividido* (figura 2b) a través de la segunda capa neuronal en un par de subvolúmenes complementarios para patrones de activación neuronal que representan varias caras y no-caras respectivamente. El primer subvolumen se divide a su vez en dos subvolúmenes mutuamente excluyentes, uno para las caras masculinas y otro para las caras femeninas. Y en cada uno de estos dos subvolúmenes hay «lugares especiales» de propietarios para cada uno de los individuos designados que la red aprendió a reconocer durante el entrenamiento.

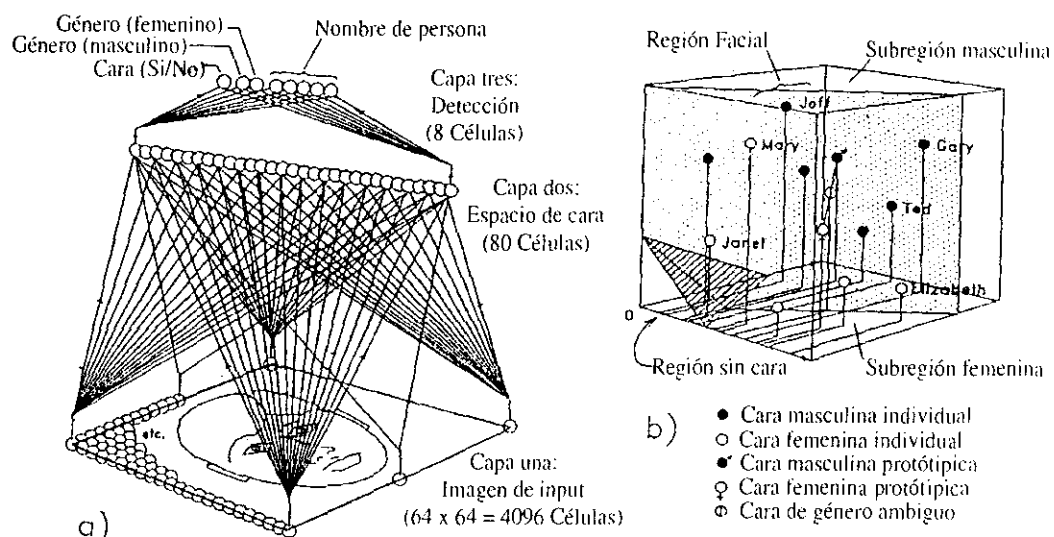


Figura 2. (a) Una red neuronal alimentada hacia adelante para el reconocimiento de rostros humanos y la distinción de sexos. (b) La jerarquía de particiones categoriales.

adquiridas durante el entrenamiento, a lo largo del espacio de patrones de activación neuronal posibles en la capa media u «oculta» de la red.

La sugerencia que avanzo aquí, siguiendo este simple modelo, es que nuestra capacidad de discriminación *moral* reside también en una intrincada matriz de conexiones sinápticas. Conexiones que dividen también, en alguna capa neuronal del cerebro humano, un espacio conceptual abstracto en un conjunto jerárquico de categorías, tales como acciones «moralmente significativas» vs. acciones «moralmente no-significativas». Y dentro de la primera categoría, acciones «moralmente malas» vs. acciones «moralmente loables». Y dentro de la primera de estas subcategorías, diversas categorías específicas, como «mentir», «engañar», «traicionar», «robar», «torturar», «asesinar», etc. (Figura 3).

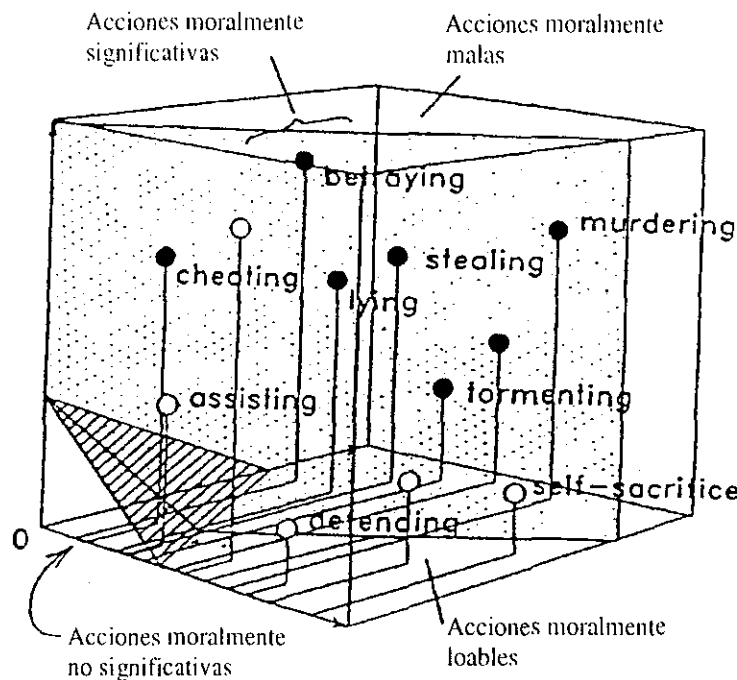


Figura 3. Un espacio de activación (conjetural) para la discriminación moral.

El espacio abstracto de patrones de activación neuronal posibles es un modelo simple de nuestro propio espacio conceptual de representaciones morales, y despliega una intrincada estructura de relaciones de semejanza y desemejanza. Estas relaciones agrupan a los vicios junto con los vicios similares y a las virtudes junto con las virtudes similares; y separan las categorías muy diferentes de acciones en sectores del es-

pacio distantes entre sí. El espacio multidimensional de similitudes (por supuesto, la Figura 3 ignora todos los ejes neuronales excepto tres) muestra una familia estructurada de «lugares especiales» categoriales o «posiciones prototipo» a las cuales son asimiladas las entradas sensoriales reales con un grado variable de cercanía.

Un espacio abstracto de patrones de activación de neuronas *motoras* proporcionará una función paralela en la génesis del comportamiento social real, una capa neuronal que gozará presumiblemente de estrechas conexiones funcionales con las neuronas sensoriales que acabamos de describir. Por todo lo dicho, estos espacios estructurados constituyen nuestro conocimiento adquirido de *la estructura del espacio social, y de cómo navegar en él.*

## 2. El aprendizaje moral

El aprendizaje moral consiste en la generación gradual de estos prototipos internos perceptivos y conductuales, un proceso que requiere práctica o exposiciones repetidas a varios *ejemplos* de las categorías perceptivas o motoras en cuestión. En una red neuronal artificial este aprendizaje consiste en el ajuste repetido de los pesos y de sus miríadas de conexiones sinápticas. Ajustes que están guiados por los sencillos *fracasos* de ejecución iniciales de la red, medidos por un programa «maestro» distintivo. En las criaturas vivientes aprender consiste también en el ajuste repetido de miríadas de conexiones sinápticas. Un proceso dirigido asimismo por las continuas experiencias de fracaso. Nuestras «tecnologías de aprendizaje» artificiales son actualmente un pobre y pálido reflejo de lo que pasa en un cerebro real, pero en ambos casos –en la red artificial y en el cerebro vivo– estos reajustes sinápticos graduales conducen a un espacio multidimensional de similitudes adecuadamente estructurado. Un espacio dividido en una familia jerárquica de subespacios categoriales, los cuales contienen un lugar especial central que representa un caso *prototípico* de esa categoría.

Habitualmente este aprendizaje lleva tiempo, y a menudo un tiempo largo. Y tal como han ilustrado los modelos de redes, este aprendizaje frecuentemente ha de estar estructurado, en el sentido de que las habilidades perceptivas o conductuales relevantes más simples han de aprenderse primero, mientras que las más complejas han de aprenderse después y sólo después de que lo hayan sido las elementales. Además, este aprendizaje puede presentar algunas patologías conocidas que derivan de un número escaso de ejemplos para el entrenamiento o de que éstos estén sesgados. En tales casos, el marco categorial debidamente adquiri-



do por la red fracasa a la hora de representar el alcance completo y la estructura verdadera del dominio social/moral que debe representar. El resultado inevitable son fallos de ejecución.

Estas observaciones apenas introducen el tema del aprendizaje moral, pero necesitamos más. El tema será tratado de nuevo más adelante, cuando discutamos el progreso moral.

### 3. La percepción moral

La más fundamental de nuestras habilidades morales consiste en la *activación*, en alguna capa apropiada de neuronas alejada al menos media docena de conexiones sinápticas de la periferia sensorial, de un patrón específico de niveles de excitación neuronal que está lo suficientemente cerca de algún patrón moral *prototipo* previamente aprendido. Este patrón de activación de la enésima capa está causado conjuntamente por el patrón de activación actual a través de una o más de las capas cerebrales sensoriales o de entrada y por las series de conexiones sinápticas cuidadosamente entrenadas que intervienen. La percepción moral forma así pieza única con la percepción en general, y su perfil presenta rasgos muy familiares a los psicólogos de la percepción.

Por ejemplo, nuestros juicios espontáneos sobre la configuración social y moral de nuestro entorno actual son enormemente sensibles a los rasgos contextuales, a la información colateral, y a nuestros intereses y focos de atención actuales. De este modo, la percepción moral está sujeta a «efectos de facilitación» y a «efectos enmascaradores». Asimismo, la percepción moral muestra la tendencia familiar de las criaturas cognitivas a «saltar a las conclusiones» en sus interpretaciones perceptivas de entradas perceptivas parciales o degradadas. Al igual que las redes artificiales, nosotros los seres humanos tenemos una fuerte tendencia a asimilar automáticamente nuestras circunstancias perceptivas actuales al más cercano de los prototipos morales disponibles que haya creado en nosotros el entrenamiento previo.

### 4. La ambigüedad moral.

Una situación es moralmente ambigua cuando es problemática en razón de su tendencia a activar *más de un* prototipo moral, prototipos que invitan a dos cursos de acción incompatibles o mutuamente excluyentes. De hecho, hasta cierto punto, la ambigüedad es un rasgo crónico de nuestra experiencia moral, en parte debido a que el mundo social es indefinidamente complejo y variado, y en parte debido a que

los intereses e informaciones colaterales que cada uno de nosotros porta en la tarea de interpretar el mundo social difieren de una persona a otra y de una ocasión a otra. Los caminos recurrentes o descendentes en el cerebro (ilustrados en la figura 1b en trazos gruesos) proporcionan una corriente continua de esta información (o desinformación) de fondo al proceso de interpretación perceptiva y activación de prototipo. Las «tomas perceptivas» diferentes de la misma situación son por tanto inevitables. Lo que nos lleva a la siguiente cuestión.

### 5. El conflicto moral.

La activación de prototipos morales distintos puede darse en dos o más individuos distintos enfrentados a la misma situación, e incluso en un sólo individuo, como cuando algún rasgo contextual es magnificado o minimizado alternativamente y nuestra toma perceptiva global oscila entre dos patrones distintos de activación en las cercanías de dos prototipos distintos. En tal caso, el individuo está en un conflicto moral («¿Debo *proteger* los sentimientos de un amigo guardando silencio acerca de una calumnia banal pero hiriente de un tercero, o debo ser franco y *veraz* en mis revelaciones a un amigo?»).

Los conflictos *interpersonales* sobre el status moral de alguna circunstancia reflejan los mismos tipos de interpretaciones divergentes, dirigidas esta vez por divergencias interpersonales en las respectivas informaciones colaterales, focos de atención, esperanzas y temores, y otros elementos contextuales que aporta cada perceptor a la situación ambigua. De este modo, los conflictos morales ocasionales son posibles, de hecho son inevitables, incluso entre individuos que poseen un entrenamiento moral idéntico y comparten categorías morales idénticas.

Finalmente, se da el caso extremo en el que los juicios morales divergen debido a que los dos individuos en conflicto poseen marcos conceptuales morales fundamentalmente diferentes, los cuales reflejan diferencias importantes en la estructura adquirida de sus respectivos espacios de activación. Aquí incluso la comunicación se vuelve difícil, y lo mismo sucede con el proceso mediante el cual se suelen resolver los conflictos morales.

### 6. Los argumentos morales.

En la imagen que estamos explorando aquí, la concepción standard del argumento moral como deducción formal de conclusiones morales a partir de premisas morales compartidas comienza a parecer extremada-

mente forzada. En su lugar, la administración y resolución de conflictos morales emerge como un proceso mucho más dialéctico por el cual los individuos en conflicto se turnan para subrayar o realzar ciertos aspectos de la situación de que se trate, y se turnan para defender diversas similitudes entre la situación y diversos prototipos compartidos, con la esperanza de producir en el adversario un patrón de activación que esté más cerca del prototipo que cada uno defiende («¡Se trata de un puñado de células sin mente, por amor de Dios! La mujer no está obligada a preservarlo o defenderlo.») y/o más lejos del prototipo que cada uno ataca («No, ¡es una persona en miniatura! Claro que está obligada.»). Se trata de empujar el punto actual de activación neuronal del interlocutor *fuera* de la categoría-atractor que la ha capturado hacia *dentro* de una categoría-atractor distinta. Se trata de intentar cambiar la probabilidad, o la fortaleza, o la proximidad a un patrón-prototipo neuronal compartido, de la conducta neuronal del oponente.

En el caso menos tratable en el cual los oponentes no comparten una familia común de prototipos morales, el argumento moral ha de tomar una forma diferente. Pospondré la discusión de esta forma más profunda de conflicto hasta la sección sobre el progreso moral.

## 7. Las virtudes morales.

Éstas son las diversas habilidades producidas en el aprendizaje social normal de *percepción social, reflexión, imaginación y razonamiento social, y navegación y manipulación social*. En la infancia se debe llegar a apreciar la multidimensional estructura de fondo del espacio social –sus oficios, sus prácticas, sus prohibiciones, su comercio– y se debe aprender a reconocer de forma rápida y fiable su configuración local. Se debe también aprender a reconocer la posición propia actual en él, y la posición a menudo muy diferente de los otros. Se debe aprender a anticipar el despliegue normal de este comercio continuo, a reconocer y a ayudar a arreglar sus patologías ocasionales, y a navegar por su estructura fluida al tiempo que se evitan los desastres sociales, grandes y pequeños. Todo esto requiere habilidad para adivinar las percepciones sociales y los intereses personales de los demás, y habilidad para manipular y negociar nuestra conducta colectiva.

Tratándose de habilidades, tales virtudes se adquieren inevitablemente de forma muy lenta, como sabe cualquiera que haya criado niños. Y su desarrollo continuado no tiene por qué acabar nunca, al menos en individuos con las oportunidades continuas y la inteligencia necesaria para refinarlas. Las estructuras adquiridas en nuestros espacios de acti-

vación neuronal –tanto perceptiva como motora– pueden ser esculpidas continuamente por la experiencia en curso y de este modo se puede aspirar a una intuición cada vez más profunda y a una comprensión efectivamente controladora de la realidad social que nos rodea. Tratándose de habilidades, son también adquiridas de forma diferente por individuos distintos, y son adquiridas de forma diferente por el mismo individuo. Cada cerebro es ligeramente distinto de todos los demás en su estructura física inicial, y la historia del aprendizaje de cada cerebro es única en una miríada de detalles. No hay dos personas idénticas en el perfil de habilidades adquiridas, lo cual nos lleva a la siguiente cuestión.

#### 8. El carácter moral.

El carácter moral único de una persona es sólo el perfil individual de sus habilidades perceptivas, reflexivas y comportamentales en el dominio social. Después de lo que se ha dicho, resultará evidente que el carácter moral se distingue más por su rica diversidad a través de los individuos que por su monotonía. Probablemente esto es algo bueno, dada la dificultad para especificar claramente un perfil canónico como ideal único. Además de la complejidad sin fin del espacio social, la existencia de una diversidad de caracteres morales simplemente refleja una tendencia saludable a explorar este espacio y a explorar los estilos más efectivos de navegación en él. Con esto no pretendo dar satisfacción a los nihilistas morales. Para hacerlo, habría que negar la realidad del aprendizaje social. Lo que estoy apoyando aquí es la idea de que se hace mejor servicio al aprendizaje moral a largo plazo en la raza humana si se tolera una distribución gaussiana de «experimentos» instruidos que si se insiste en una ortodoxia estrecha e imposible.

Esta concepción del conjunto de las virtudes morales como una red de habilidades lentamente adquiridas contiene también una crítica implícita de un absurdo romanticismo popular, a saber, la idea de una «conversión repentina» a la moralidad, tal como aparece ejemplificada por la «cara lacrimosa del pecador arrepentido» y por el cristiano carismático que «renace» tras el bautismo. El carácter moral no es algo –ni siquiera remotamente– que pueda adquirirse en un día por un Acto de Voluntad o por una sola Visión.

La idea de que puede ser adquirido de este modo constituye una reflexión refutadora de dos concepciones familiares del carácter moral que con lo dicho quedan desacreditadas. La primera identifica el carácter moral con la aceptación de un conjunto canónico de reglas para orientar la conducta. La segunda identifica el carácter moral con un conjunto

canónico de deseos, como el deseo de maximizar la felicidad general, etc. Quizás se pueda abrazar un conjunto de reglas en un acto catártico, y quizás se pueda privilegiar permanentemente algún conjunto de deseos por un acto de voluntad. Pero ningún acto puede dar como resultado lo que se necesita realmente, a saber, un conjunto intrincado de habilidades perceptivas, reflexivas y sociomotoras bien afiladas. Adquirirlo lleva varias décadas. Las epifanías de compromiso moral pueden señalar, a lo sumo, el inicio de este proceso. Por supuesto que las iniciaciones son bienvenidas, pero no le damos un diploma de educación secundaria a un niño que acude el primer día a la escuela primaria. Por las mismas razones, los caracteres morales «renacidos» deben esperar probablemente un periodo similar de tiempo antes de celebrar su realización moral o de imponer su autoridad moral.

### 9. La patología moral.

Este es un tema extenso, dado que si hay muchas formas diferentes de lograr ser una criatura moralmente madura, hay más formas aún de no lograrlo. Para empezar, la patología moral consiste en la ausencia parcial, o corrupción subsiguiente, de la constelación normal de habilidades perceptivas, reflexivas y comportamentales acerca de las cuales estamos hablando. En términos de la teoría cognitiva que subyace al presente enfoque, consiste en el fracaso para alcanzar, o para activar normalmente, una jerarquía adecuada de prototipos morales en el espacio de activación neuronal de una persona. Y en el nivel más bajo, esto consiste en un fracaso, temprano o tardío, para lograr y mantener la configuración conveniente de los  $10^{14}$  pesos sinápticos del cerebro, configuración que sustenta la jerarquía deseada de prototipos y hace posible su activación apropiada.

Los términos «normalmente», «adecuado», «conveniente», «apropiado» aparecen en esta rápida caracterización, y todos deben su significado a una mezcla inextricable de comprensión *funcional* en el ámbito de la neurobiología cognitiva y de auténtica comprensión *moral*, como la que se encuentra en el sentido común y en el código civil y penal. Lo importante aquí es que entendamos que las manifestaciones de incompetencia moral, tanto las grandes como las pequeñas, son a menudo el reflejo de fallos funcionales específicos, tanto grandes como pequeños, en el cerebro. Esto no es una observación especulativa. Gracias a la creciente disponibilidad de las tecnologías de exploración cerebral, tal como la Tomografía de Emisión de Positrones (TEP) y la Imagen de Resonancia Magnética (IRM), los neurólogos se han familiarizado con una va-

riedad de formas muy específicas de daños cerebrales que se presentan en sintonía con fallos cognitivos en la percepción moral, el razonamiento moral y la conducta social (Damasio et al., 1991; Damasio, 1994; Bechara et al., 1994; Adolphs et al., 1996).

Dos ejemplos rápidos ilustrarán la cuestión. Los neurólogos Antonio y Hanna Damasio tienen un paciente, conocido en la literatura como «Boswell», que se hizo famoso independientemente por su incapacidad para asentar cualquier nuevo recuerdo a largo plazo debido a lesiones bilaterales en su lóbulo temporal medio, incluyendo el hipocampo. A causa de su enfermedad, su «pasado recordado» es una ventana móvil que no alcanza más que los últimos cuarenta segundos. Y lo que es más importante para nuestros propósitos, más tarde se vio que también muestra una curiosa incapacidad para «ver el mal» en imágenes con escenas emocionalmente cargadas y potencialmente violentas. En particular, es incapaz de captar diversas emociones negativas expresadas en las *caras* de las personas, y discursará alegremente sobre explicaciones inocentes de escenas social y moralmente problemáticas que se le muestren. Sin embargo, no hay nada que esté mal en los ojos o en el sistema visual de Boswell. Su deficiencia cognitiva reside en una docena aproximadamente de pasos sinápticos y en una docena de capas neuronales detrás de sus retinas.

Tal como reveló el examen IRM, la encefalitis herpes-simplex de Boswell había dañado también la mitad inferior de sus lóbulos temporales, que incluyen el área llamada «IT» (infero-temporal), conocida por su papel fundamental en la discriminación de rostros humanos y en la codificación de expresiones faciales. Ya no podía reconocer la identidad de caras que conocía muy bien antes de su enfermedad (estrellas de cine y presidentes, por ejemplo), y su percepción moral se había deteriorado selectivamente de la forma descrita.

Un segundo paciente, EVR, llevaba una vida normal como respetado contable, padre devoto y marido fiel. Cuando estaba en torno a los cuarenta y cinco años, se le extirpó con éxito un tumor cerebral en la región frontal ventromedial, y los tests posteriores no revelaron ningún cambio en su coeficiente de inteligencia de 140. Pero a los seis meses había perdido el trabajo por su desenfrenada irresponsabilidad, había tomado una serie de decisiones financieras perjudiciales, se había divorciado de su contrariada esposa, había estado casado brevemente con una prostituta que luego le había abandonado, y se había vuelto incapaz de tener la prudencia normal que guía la elaboración de planes complejos y las interacciones sociales intrincadas. Los exámenes IRM posteriores confirmaron que la intervención quirúrgica del tumor había lesionado el

córtex frontal ventromedial (la sede de la planificación compleja) y sus conexiones con la amígdala (una área límbica primitiva que aparentemente regula el miedo y la ansiedad).

La consecuencia funcional de esta ruptura en el cableado fue la de *aislar* el razonamiento práctico de EVR de las reacciones somáticas «viscerales» y emocionales que acompañan normalmente a la evaluación racional de las alternativas prácticas. En las personas normales, estos «marcadores somáticos» (como los han denominado los Damasio) constituyen una dimensión importante de la información socialmente relevante y un factor clave en la inhibición de las decisiones. En EVR habían sido desconectados del circuito, dando como resultado la conducta que hemos descrito.

Estos dos fallos, el de la percepción moral y el de la conducta moral respectivamente, fueron el resultado de una enfermedad repentina y el consecuente daño de áreas cerebrales específicas, lo cual atrajo la atención de la profesión médica y condujo a su examen detallado. Pero éstos y muchos otros déficits neuronales pueden aparecer también lentamente, como resultado de un desarrollo malaventurado y otras enfermedades crónicas –infecciones infantiles, toxinas de bajo nivel, metabolismo anormal, química cerebral anormal, nutrición anormal, uso materno de drogas durante el embarazo, etc. Permítaseme subrayar que con esto no estoy sugiriendo que todos los fallos del carácter moral puedan ser achacados a déficits estructurales en el cerebro. Una educación moral adecuada –esto es, un periodo largo de socialización intrincada– sigue siendo una condición necesaria para adquirir un carácter moral bien formado, e indudablemente, la gran mayoría de los fallos, sobre todo los pequeños, pueden ser achacados por completo a diversas carencias en este proceso.

Aún así, el proceso educativo está enteramente entrelazado con el proceso de desarrollo y es profundamente dependiente de la existencia de estructuras cerebrales normales que encarnen la matriz deseada de habilidades. Por tanto, al menos algunos fallos del carácter moral, y especialmente los más graves, implican probablemente alguna discapacidad o marginalidad en el nivel de la estructura cerebral y/o en la actividad fisiológica. Si queremos tratar con acierto en las leyes y en el sistema correccional estos grandes fallos en el carácter moral, haríamos bien en entender las múltiples dimensiones de fallos neuronales que colectivamente pueden originarlos. No podemos arreglar lo que no entendemos.

## 10. La reeducación moral.

Consideremos primero el cerebro estructural y fisiológicamente *normal* cuyo entorno social formativo no consigue proporcionar una educación moral normal. La experiencia del niño puede carecer de ejemplos cotidianos de comportamiento moral normal en los demás, puede carecer de las oportunidades para participar en prácticas sociales normales, puede no conseguir observar cómo los demás tratan con éxito y de forma rutinaria los conflictos sociales inevitables, y puede carecer del historial normal de reeducación de su percepción y su conducta por parte de los hermanos mayores o de los padres. Para el joven problemático que resulte de ello, la reeducación moral consistirá obviamente en intentar compensar de algún modo una educación errada o insatisfactoria.

Eso puede resultar muy difícil. La plasticidad cognitiva y el afán por imitar que encontramos en los niños se reduce considerablemente en el adulto joven. Y un adulto joven no puede encontrar fácilmente el tipo de comunidad tolerante de compañeros inocentes y mayores juiciosos en la que la mayor parte de los niños tienen la fortuna de crecer. De este modo pierden no una, sino dos oportunidades importantes.

El problema se complica por el hecho de que en un ambiente social empobrecido como el descrito no es que los niños simplemente no consigan aprender. Al contrario, pueden aprender muy bien, pero *lo que* aprenden es un conjunto completamente pervertido de prototipos sociales y morales, acompañados de una familia de habilidades que –aunque quizás sean funcionales en el entorno empobrecido que las conforma– son decididamente *disfuncionales* en la estructura más coherente de la sociedad en general. Esto significa que el adulto joven debe efectuar un *desaprendizaje* sustancial. Y dada la gran «inercia» cognitiva que caracteriza incluso a los seres humanos normales, la pendiente correctora es más escarpada, especialmente cuando el delincuente joven es encerrado en una comunidad carcelaria fuertemente unida de agentes sociales igualmente pervertidos.

Se supone que este artículo no pretende recomendar ninguna política social o moral, pero aquéllos que se dedican a estas cuestiones pueden encontrar relevantes los siguientes hechos. Se dice que el presupuesto que América dedica a las prisiones estatales y federales es algo mayor que el presupuesto que dedica a *toda* la educación superior, contando sus universidades con investigación de élite, sus grandes universidades estatales, sus miríadas de *colleges* de artes liberales, todas sus escuelas técnicas y sus *junior colleges* bianuales. Es concebible al menos que nuestro enorme presupuesto penitenciario podría ser gastado



más sabiamente en políticas preventivas que tuvieran como objetivo aumentar la calidad del entorno social de los niños desfavorecidos, en lugar de gastarlo en políticas que se esfuerzan, con todos los factores en contra, por reparar los daños una vez que ya están hechos.

Por supuesto que la opción no es un cambio convulsivo. Sea lo que sea lo que nuestras prisiones hacen o dejan de hacer, mantienen al menos fuera de las calles y de nuestro comercio social a algunos de los agentes sociales más peligrosamente incompetentes y a los declaradamente dañinos. Sin embargo, la plasticidad del joven frente a la persona mayor plantea una constante invitación a cambiar nuestros recursos correctivos en la tutela de los niños, tal como dicta la debida prudencia. Esta sugerencia política aspira a reducir el número absoluto de entradas en nuestras instituciones correccionales. Pero un asunto igualmente importante sobre el que ahora volveré es cómo incrementar, anticipándonos a estos avances «utópicos», la proporción de salidas.

En tal sentido haré una observación final sobre las personas normales. La plasticidad cognitiva de un joven—esto es, su capacidad inigualable para aprender— se debe a factores neuroquímicos y fisiológicos que se desvanecen con la edad. (Según una teoría, y hay otras, algunas conexiones sinápticas son modificadas selectivamente por la producción local y la difusión de óxido nítrico en el cerebro.) Supongamos que pudiéramos aprender cómo *recrear* en un adulto joven, de forma temporal y con medios neurofarmacológicos, el régimen completamente normal de plasticidad neuronal y aptitud para el aprendizaje que tienen los niños. En conjunción con algunos programas más efectivos de resocialización que los que actualmente disponemos (sin ellos la farmacología sería una pérdida de tiempo), esto podría relanzar a las «personas normales desaventajadas» a algo mucho más próximo a una trayectoria social normal y fuera de las cárceles para siempre.

Pero quedan todavía los anormales genuinos, cuya reeducación moral consiste ante todo en intentar reparar o compensar algún(os) defecto(s) estructural o fisiológico en las funciones cerebrales. Incluso si estas personas fuesen casos perdidos y sólo cupiera mantenerlos encarcelados permanentemente o fuera de la corriente social de algún otro modo, sería útil a la política social identificarlos de forma fiable. Pero algunos al menos no serán casos perdidos. Allá donde el déficit sea de naturaleza bioquímica—generando, por ejemplo, perfiles emocionales crónicamente inapropiados—, la intervención neurofarmacológica, en la forma ya familiar de implantes crónicos subdurales, hará quizás que algunas víctimas puedan volver a algo parecido a una economía neuronal normal y a un perfil emocional normal. Esto ya sería un beneficio suficiente, pero

además estas personas serían también candidatos para las técnicas de resocialización que exponíamos antes para las personas normales des-aventajadas.

Esta discusión presupone una comprensión neurológica mayor de la que poseemos, y resulta claramente especulativa. Pero permite ilustrar algunas direcciones en las que muy bien podríamos desear movernos una vez que nuestra comprensión haya madurado. En cualquier caso, cerraré esta discusión volviendo a subrayar la importancia universal de la socialización gradual mediante una larga interacción con un orden moral previamente existente. Nunca podremos crear un carácter moral mediante la sola intervención médica. Hay demasiados billones de conexiones sinápticas para que se les pueda asignar los pesos apropiados, y sólo una larga experiencia puede aspirar a realizar esta tarea superlativamente complicada. El objetivo de explorar las tecnologías mencionadas será maximizar las oportunidades de todos para participar y beneficiarse de estos procesos tradicionales e irremplazables.

### 11. La diversidad moral.

Con esto no me refiero a la curva multidimensional de campana de la diversidad de los caracteres morales en una cultura y una época dadas, sino a la no-identidad en dos culturas separadas en el espacio y/o el tiempo de todo el *sistema* común a la mayor parte de los miembros normales de cada una de prototipos morales y habilidades apreciadas. Estas diferencias importantes en la conciencia moral reflejan típicamente diferencias en las circunstancias económicas sustantivas entre las dos culturas, en las amenazas peculiares al orden social a las que tienen que hacer frente, en las tecnologías de las que disponen, en las creencias metafísicas que mantienen, y en otros accidentes de la historia.

Cuando se descubre tal diversidad, se la ve a menudo como una base para un escepticismo general sobre la objetividad o realidad del conocimiento moral. Este fue ciertamente el efecto que tuvo sobre mí al final de mi infancia, una reacción que se vio reforzada por el sorprendentemente bajo nivel de argumentación moral que oíría en mis compañeros de escuela más religiosos, e incluso desde los púlpitos locales. Pero ya no es ésta mi reacción, puesto que a lo largo de la historia ha habido diferencias comparables entre distintas culturas en lo concerniente al conocimiento *científico*, y una estupidez semejante en el razonamiento puramente «factual» (piénsese en la «medicina New Age» o en la «UFOlogía», por ejemplo). Y sin embargo esta diversidad real y esta falta de rigor igualmente lamentable no respalda un escepticismo

general sobre la posibilidad del conocimiento científico. Sólo muestra que no es fácil adquirirlo, y que su logro requiere un largo proceso de evaluación cuidadosa y honesta de una amplia variedad de experimentos complejos sobre un gama sustancial de experiencias humanas. Lo que señala a nuestro siguiente asunto.

## 12. El progreso moral.

Si existe el progreso moral –hay alguna disputa sobre ello–, éste consiste en el lento cambio y desarrollo, a través de periodos históricos, de los prototipos morales que enseñamos a nuestros niños e imponemos por la fuerza a los adultos marginales, un proceso de desarrollo que es instruido gradualmente por nuestra *experiencia* colectiva de una vida colectiva vivida bajo estos prototipos conformadores de la percepción y orientadores de la conducta.

Desde una perspectiva neurocomputacional, este proceso sólo se muestra diferente en su enfoque ontológico de lo que nos complacemos en llamar *progreso científico* –el mundo *social* en tanto que opuesto al mundo *natural*. También en las ciencias naturales, para lograr la competencia de un adulto, hay que adquirir una compleja familia de habilidades perceptivas, reflexivas y comportamentales en el campo relevante. Y también allí estas habilidades están encarnadas en un conjunto adquirido de prototipos estructurales, dinámicos y manipulacionales. Al contrario de lo que ocasionalmente dicen algunas voces deflacionistas, nuestro progreso científico a lo largo de los siglos es una realidad dramática y estimulante, y es en parte el resultado de una miriada de enseñanzas (frecuentemente penosas) de una vida continuada de experimentación y tecnología vivida bajo estos mismos prototipos científicos conformadores de la percepción y orientadores de la conducta.

Sugiero que nuestro desarrollo conceptual en el ámbito moral difiere sólo en los detalles de nuestro desarrollo en el ámbito científico. Incluso contamos con instituciones cuya función es afinar continuamente y ocasionalmente rehacer nuestras concepciones acerca de la conducta adecuada, la práctica permisible y el comportamiento proscrito. Los cuerpos legislativos municipal, estatal y federal vienen inmediatamente a la cabeza, así como el cuerpo de funcionarios y los diferentes niveles de la administración de justicia, con su conjunto en permanente evolución de casos y precedentes legales orientadores de las decisiones. Al igual que sucede con nuestras instituciones en las ciencias empíricas, estas instituciones socialmente orientadas suelen sobrevivir siglos, y a veces muchos siglos, a la gente que pasa por sus oficinas. Y al igual que sucede

con los resultados de nuestras instituciones científicas, el resultado aquí es la acumulación de niveles sin precedentes de experiencia (social) registrada, así como los beneficios equilibradores de la toma colectiva de decisiones, y el logro de niveles de comprensión moral que son inalcanzables para un sólo individuo durante su vida.

Se puede objetar a este paralelismo general que la ciencia se dirige a la naturaleza última de una realidad fija, estable e independiente, en tanto que nuestras instituciones sociales, legislativas y legales se dirigen a una realidad plástica que depende profundamente de la actividad organizadora de los seres humanos. Pero este supuesto contraste desaparece casi por completo cuando se contemplan tanto las adquisiciones de la ciencia como las de la sabiduría moral como adquisiciones de un conjunto de *habilidades*. Ambas se dirigen a una parte supuestamente *no-plástica* de sus respectivos dominios –las leyes básicas de la naturaleza en el primer caso y la naturaleza humana básica en el segundo. Y ambas se dirigen a una parte profundamente *plástica* de sus respectivos dominios –la articulación, manipulación y explotación tecnológica del mundo natural en el caso de la ciencia en acción, y la articulación, manipulación y explotación práctica de la naturaleza humana en el caso de las morales y las políticas en acción. Una ciudad próspera representa simultáneamente el éxito en ambas dimensiones de la actividad cognitiva humana. Y las tecnologías artificiales que resultan, tanto las naturales como las sociales, posibilitan una visión más profunda del carácter básico del universo natural y de la naturaleza humana respectivamente.

### 13. La unidad/sistematicidad moral.

Este paralelismo con las ciencias naturales tiene dimensiones adicionales. Así como el progreso en la ciencia conduce de vez en cuando a una grata unificación de nuestra comprensión –como cuando se ve a todos los movimientos planetarios como casos especiales del movimiento de proyectiles, y a todos los fenómenos ópticos como casos especiales de ondas electromagnéticas–, así también el progreso en la teoría moral conduce a intentos ocasionales de unificación conceptual –como cuando se presenta el conjunto de nuestras obligaciones y prohibiciones como elementos de un *contrato social* (Hobbes), o como instanciaciones locales de un *imperativo categórico* (Kant), o como el reflejo de *reglas elegidas racionalmente tras un velo de ignorancia personal* (Rawls). Estas y otras sugerencias familiares son intentos rivales de unificar y sistematizar nuestras intuiciones morales dispersas o nuestra comprensión moral previa, y comportan (o se espera que comporten) el mismo tipo de

virtudes desplegadas por las reducciones interteóricas en la ciencia, a saber, mayor simplicidad en las concepciones unificadas, mayor consistencia en su aplicación, y una capacidad intensificada (proveniente de un incremento en la generalidad) para tratar nuevos tipos de problemas sociales y morales.

Como sucedía con los aspectos anteriores de la cognición moral, también este tipo de logro cognitivo a gran escala es comprensible en términos neurocomputacionales, y parece implicar los mismos tipos de cambios neurodinámicos que (se supone que) están implicados cuando se dan intuiciones teóricas en las ciencias naturales. Más específicamente, hay un amplio elenco de fenómenos perceptivos –los cuales (supongamos) suelen activar una gran cantidad de prototipos morales,  $m_1, m_2, m_3, \dots, m_n$ – que se procesan bajo un nuevo régimen de manipulación recurrente (recuérdense los caminos neuronales recurrentes de la figura 1b), y ello da como resultado la activación de un prototipo moral inesperado  $M$ . Un prototipo cuyo despliegue típico ha estado por completo hasta entonces en otros dominios perceptivos. Un prototipo que emerge ahora como un prototipo *superordenado*, del cual los prototipos menores dispersos  $m_1, m_2, m_3, \dots, m_n$  pueden verse retrospectivamente como ejemplos subordinados.

Lo anterior es una descripción en términos de redes neuronales de lo que sucede cuando, por ejemplo, se *axiomatiza* nuestro conocimiento disperso acerca de algún área. Pero la axiomatización, en un lenguaje formal al modo en que aparece en los libros de texto, no es sino un caso menor de este proceso mucho más general. Un proceso que abarca también las diversas formas del conocimiento *no* discursivo. Un proceso que abarca tanto a la ciencia como a la ética.

#### 14. Reflexiones acerca de algunas recientes «éticas de la virtud».

Como habrán visto la mayor parte de los filósofos, el retrato general del conocimiento moral que emerge de los modelos de cognición en términos de redes neuronales es un retrato que ya está sometido a un examen activo en la filosofía moral, con independencia de cualquier conexión que pueda haber con la neurobiología cognitiva. Su defensor original es Aristóteles y la comunidad de sus investigadores actuales incluye figuras tan intelectualmente distintas como Mark Johnson (1993), Owen Flanagan (1991), y Alasdair MacIntyre (1981), cada uno de los cuales llegó a esta perspectiva general por razones enteramente propias. Por las razones esbozadas en este artículo, estoy obligado a contarme entre ellos (de lo cual me honro). Pero no me encuentro completamente

a gusto en este grupo, ya que dos de los filósofos que he mencionado tienen una visión del progreso moral muy diferente de la que acabo de esbozar. Flanagan (1996) ha expresado sus francas dudas de que la conciencia moral humana realice alguna vez un «progreso» genuino, y sugiere que sus cambios ocasionales pueden ser vistos más adecuadamente como movimientos sin dirección efectuados como respuesta puntual a los cambios en nuestro entorno económico y social.

MacIntyre (1981) ofrece una visión diferente pero igualmente escéptica. En ella añora la inocencia perdida de las comunidades humanas anteriores a la Ilustración, que estarían mucho más estrechamente unidas por un tejido compacto de prácticas sociales compartidas, las cuales proporcionarían la clase de entorno altamente interactivo y de mutua dependencia necesario para que muchas virtudes morales se desarrollen y florezcan. Lamenta la emergencia de la vida social posterior a la Ilustración, liberal, secular y comparativamente anónima e independiente, dirigida por los hombres modernos industriales, pues, según nos dice, el rico suelo necesario para el aprendizaje moral se ha empobrecido. Las virtudes morales familiares deben ser ahora adquiridas, pulidas y ejercitadas en lo que es, comparativamente, un vacío social. Si hemos experimentado algo en los últimos siglos ha sido un *regreso moral*.

Estoy en desacuerdo con ambos autores, y seré minucioso explicando por qué. Para empezar con MacIntyre, acepto de entrada su crítica de la concepción caricaturesca de la Ilustración (británica) del *homo economicus*, un calculador hedonista casi completamente desprovisto de todo interés en evaluar los deseos mismos que dirigen sus cálculos, o bien de recursos para hacerlo. Asimismo acepto su crítica de la concepción de la Ilustración (continental) de la *razón pura* como la clave para identificar un conjunto único de reglas orientadoras de la conducta. Y estas concesiones no las hago de mala gana. Coincido de todo corazón con MacIntyre en que ninguna de las dos concepciones arroja mucha luz sobre la naturaleza de la virtud moral.

Pero por burdas que sean estas ideas morales o metamorales, están todavía lejos de los conceptos aún más caricaturescos de *homo ovejicus* y *homo infanticus* que la iglesia cristiana preilustrada propugnaba implacablemente. Retratar a la humanidad como ovejas guiadas por un Pastor sobrenatural, o como niños agradecidos a un Padre sobrenatural, era un autoengaño incluso más oscuro y era incluso menos probable que sirviera como medio para ascender en el camino de la comprensión moral.

Puedo estar equivocado en estas terminantes afirmaciones, y si lo estoy, que así sea. Pues lo dicho en el anterior párrafo *no* expresa el argumento verdaderamente importante en favor del progreso moral de-

bido a la Ilustración. Ese argumento descansa en otra parte. Descansa en la permanente apertura de una tradición de cuidadosa *tolerancia* de una diversidad de comunidades locales, cada una de ellas unida por su propia fábrica de prácticas sociales. Descansa en el establecimiento de instituciones perdurables para la *evaluación* basada en fuertes principios de los diversos modos de organización social y para la *crítica* institucionalizada de algunos de ellos y la *emulación* sistemática de otros. Descansa en suma en el hecho de que la Ilustración acabó con el dominio de una calcificada dictadura moral y la sustituyó por una tradición que estuvo finalmente dispuesta a aprender de su deliberadamente amplia experiencia y de sus inevitables errores en las normas morales de primer orden.

De nuevo voy a recurrir a un paralelismo sobresaliente. La virtud de la Ilustración en la esfera moral fue precisamente la misma virtud que desplegó en la esfera científica, a saber, la legitimación de la diversidad teórica responsable y el establecimiento de instituciones perdurables para su evaluación crítica y explotación positiva. Es este proceso a largo plazo, y no una teoría moral o una práctica moral particular que pudiera haber ocupado fugazmente su atención, lo que marca el logro principal de la Ilustración.

MacIntyre comienza su Introducción de *After Virtue* con un escenario intelectualmente estimulante de ciencia ficción en el que se describe la pérdida de una compleja tradición práctica que por si sola da vida a la correspondiente familia de términos teóricos, y la relativa esterilidad de su uso continuado en ausencia de la tradición que los sustenta. Esto expresa lo esencial de su crítica a nuestra historia moral desde la Ilustración. Pero, para una evaluación crítica, podemos construir fácilmente una crítica paralela de nuestra historia *científica* en el mismo periodo, y sugiero que este paralelismo arroja alguna luz sobre la perspectiva bastante conservadora de MacIntyre.

Consideremos el apogeo de la Ciencia Aristotélica, desde el siglo cuarto a. de C. al siglo diecisiete de nuestra era (un periodo más largo incluso que el de la dominación cristiana en el esfera moral), y consideremos el conjunto unificado y unificador de prácticas intelectuales y tecnológicas que sustentaba. Está la tradición médica que iba desde Galeno en Roma hasta los cuatro humores de los médicos medievales tardíos. Está la tradición astronómico/astrológica que se extiende desde Ptolomeo en Alejandría hasta Johannes Kepler en Praga, quien todavía realizaba horóscopos para los ricos a pesar de su teorizar apóstata. Está el conjunto complejo de prácticas industriales que mantuvieron los alquimistas desde los griegos alejandrinos hasta la Europa del siglo dieci-

siete; tradición a la que pertenecieron las importantísimas prácticas de la metalurgia y la metalistería, así como las de la tintorería y la manufactura medicinal. Estas tres tradiciones, y otras que debo pasar por alto por razones de espacio, estuvieron fuertemente ligadas por la práctica cotidiana y por su origen conceptual, y constituyeron un entorno consistente y coherente en el que podían florecer las virtudes prácticas y tecnológicas de la antigüedad tardía. Y puesto que lo hicieron, se cumple la primera condición de MacIntyre.

Y la segunda también, puesto que este unificado «paraíso» está perdido y bien perdido, habiendo sido desplazado por un avispero de ciencias diferentes, ciencias tan diversas como la astrofísica, la biología molecular, la antropología, la ingeniería eléctrica, la física del estado sólido, la inmunología y la meteorología termodinámica. La ciencia moderna ahora se dirige y avanza por tantos frentes que la práctica investigadora de los científicos individuales y la práctica tecnológica de los ingenieros individuales se aísla cada vez más de todo lo que no sean los miembros más inmediatos de las comunidades cognitivas particulares. Y las virtudes cognitivas que manifiestan están igualmente fragmentadas. Pueden incluso encontrar dificultades para comunicarse los unos con los otros.

Ya ven dónde quiero ir a parar. El florecimiento sin precedentes de muchas ciencias modernas pudo haber originado problemas –problemas reales–, pero la pérdida de una «edad dorada» anterior y más saludable no fue ciertamente uno de ellos. Aunque reales, estos problemas son simplemente el precio que la humanidad paga por su desarrollo, y ya intentamos resolverlos por medio de curricula interdisciplinarios, congresos y antologías interdisciplinarios, y por la búsqueda incesante de unificaciones explicativas y reducciones interteóricas.

Propongo, en relación con la reflexión de MacIntyre, una tesis paralela sobre nuestras instituciones morales, políticas y legales desde la Ilustración. Sin duda hay problemas que surgen del florecimiento sin precedentes de las diversas sociedades modernas industriales y de sus sub-sociedades, pero obviamente no es uno de ellos la pérdida de contacto con una edad dorada anterior. Los problemas reales planteados por la diversidad moral y política son simplemente el precio que la humanidad paga por su desarrollo. Y como en el caso de las ciencias dispersas, ya intentamos resolverlos mediante constantes remiendos legales, mediante la evolución, dirigida por la propia realidad, de los precedentes judiciales, mediante la tolerancia de «divorcios» políticos ocasionales (e. g., Yugoslavia, la Unión Soviética, el Parlamento Escocés), y mediante la búsqueda incesante de unificaciones legales, políticas y econó-



micas. Junto con el descubrimiento del fuego y el ejemplo multidoctrinal de la antigua Grecia, la Ilustración puede ser la mejor cosa que nos ha pasado nunca.

El análogo doctrinal del comunitarismo implícito de MacIntyre en teoría moral es una forma hiperbólica de conservadurismo kuhniano en filosofía de la ciencia. Un conservadurismo que valora las virtudes (muy reales) de una tradición determinada de «ciencia normal» (como la astronomía ptolemaica, la termodinámica clásica o la mecánica newtoniana) por encima de las instituciones comparativamente frágiles de evaluación colectiva, de comparación y de crítica que podrían sacar lentamente a la luz del sol sus vicios ocultos y preparar el terreno para su derrocamiento legítimo a manos de unos modos más prometedores de organización cognitiva. Uno puede ver ciertamente la razón básica del «comunitarismo» de Kuhn: las prácticas científicas estables posibilitan muchas cosas valiosas. Pero instituciones tolerantes para la evaluación y modificación de estas prácticas posibilitan cosas todavía *más* valiosas –las más obvias serían prácticas nuevas y más estables.

Esta particular defensa de la Ilustración también sustenta mi respuesta a la modalidad bastante diferente de escepticismo de Flanagan. Puesto que contemplo las cuestiones desde la perspectiva de las redes neuronales explicada anteriormente en este artículo, no puedo encontrar ninguna diferencia entre los presuntos mecanismos cerebrales y los procesos cognitivos que garantizan la cognición moral y la cognición científica. Ni puedo encontrar ninguna diferencia significativa entre las respectivas instituciones sociales que administran respectivamente nuestra conciencia científica y moral en despliegue. En ambos casos, aprender de la experiencia es el resultado perfectamente normal tanto de la maquinaria neuronal como de la social. Esto significa que el progreso moral ni es menos posible ni menos probable que el progreso científico. Y puesto que, en este momento, a ninguno de nosotros se nos hace pasar por los instrumentos de tortura de los sótanos del Vaticano, creo que este progreso es también real.

Queda la cuestión residual de si las *ciencias* realizan un progreso auténtico, pero esta cuestión la dejo para otra ocasión. Las afirmaciones netas de este artículo son que, 1) cualquiera que sea su status último, la cognición moral y la cognición científica están en pie de igualdad, puesto que usan los mismos mecanismos neuronales, mostrando el mismo perfil dinámico, y responden a presiones empíricas similares tanto a corto plazo como a largo plazo; y 2) tanto en el aprendizaje moral como en el científico, el logro cognitivo fundamental es la adquisición de *habilidades*, encarnada en la puesta a punto de la configuración de las  $10^{14}$  conexiones sinápticas cerebrales.

## REFERENCIAS

- Adolphs, R., Tranel, D., Bechara, A., Damasio, H., y Damasio, A. R.: 1996, «Neuropsychological Approaches to Reasoning and Decision Making», en A. R. Damasio et al. (eds.), *The Neurobiology of Decision-Making*, Berlin: Springer-Verlag, pp. 157-180.
- Bechara, A., Damasio, A., Damasio, H., y Anderson, S. W.: 1994, «Insensitivity to Future Consequences Following Damage to Human Prefrontal Cortex», *Cognition* 50, 1-15.
- Churchland, P. M.: 1989, *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge: The MIT Press.
- Churchland, P. M.: 1989a, «On the Nature of Theories: A Neurocomputational Perspective», in W. Savage (ed.), *Scientific Theories: Minnesota Studies in the Philosophy of Science*, Vol. XIV, Minneapolis: University of Minnesota Press, pp. 59-101. Capítulo 9 de Churchland, P. M. 1989.
- Churchland, P. M.: 1989b, «On the Nature of Explanation: A PDP Approach». Capítulo 10 de Churchland, P. M. 1989. Reimpreso en J. Misiek (ed.), *Rationality*, Vol. 175 de *Boston Studies in the Philosophy of Science*, Dordrecht: Kluwer, 1995.
- Churchland, P. M.: 1989c, «Learning and Conceptual Change». Capítulo 11 de Churchland, P. M. 1989.
- Churchland, P. M.: 1995, *The Engine of Reason, The Seat of the Soul: A Philosophical Journey into the Brain*, Cambridge: The MIT Press.
- Cottrell, G.: 1991, «Extracting Features from Faces Using Compression Networks: Face, Identity, Emotions and Gender Recognition Using Holons», en D. Touretzky, J. Elman, T. Sejnowski, y G. Hinton (eds.), *Connectionist Models: Proceedings of the 1990 Summer School*, Morgan Kaufmann: San Mateo, CA.
- Damasio, A. R.: 1994, *Descartes' Error*, New York: Putnam & Sons.
- Damasio, A. R., Tranel, D., y Damasio, H.: 1991, «Somatic Markers and the Guidance of Behavior», en H. Levin et al. (eds.), *Frontal Lobe Function and Dysfunction*, New York: Oxford University Press.
- Elman, J. L.: 1992, «Grammatical Structure and Distributed Representations», en S. Davis (ed.), *Connectionism: Theory and Practice*, Vol. 3 de la serie Vancouver Studies in Cognitive Science, Oxford: Oxford University Press.
- Flanagan, O.: 1991, *Varieties of Moral Personality: Ethics and Psychological Realism*, Cambridge: Harvard University Press.
- Flanagan O.: 1996, «The Moral Network», en B. McCauley (ed.), *The Churchlands and Their Critics*, Cambridge, Mass.: Blackwell, pp. 192-215.

- Gorman, R. P. and Sejnowski, T. J.: 1988, «Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets», *Neural Networks* 1, 75-89.
- Johnson, M.: 1993, *Moral Imagination*, Chicago: Chicago University Press.
- Lehky, S. y Sejnowski, T. J.: 1988, «Network Model of Shape-from-Shading: Neuronal Function Arises from Both Receptive and Projective Fields», *Nature* 333, 452-454.
- Lehky, S. y Sejnowski, T. J.: 1990. «Neural Network Model of Visual Cortex for Determining Surface Curvature from Images of Shaded Surfaces», *Proceedings of the Royal Society of London B* 240, 251-278.
- Lockery, S. R., Fang, Y., y Sejnowski, T. J.: 1991, «A Dynamical Neural Network Model of Sensorimotor Transformation in the Leech», *Neural Computation* 2, 274-282.
- MacIntyre, A.: 1981, *After Virtue*, Notre Dame: University of Notre Dame Press.
- Rosenberg, C. R. y Sejnowski, T. J.: 1987, «Parallel Networks that Learn to Pronounce English Text», *Complex Systems* 1, 145-168.
- Saver, J. L. y Damasio, A. R.: 1991, «Preserved Access and Processing of Social Knowledge in a Patient with Acquired Sociopathy due to Ventromedial Frontal Damage», *Neuropsychologia* 29, 1241-1249.