



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

FACULTAD DE PSICOLOGÍA

REINTERPRETANDO LA RELACIÓN ENTRE LENGUAJE, COGNICIÓN Y
COMPORTAMIENTO: EL CONOCIMIENTO PROPOSICIONAL COMO MECANISMO
COMUNICATIVO

T E S I S
QUE PARA OBTENER EL GRADO DE:
LICENCIADO EN PSICOLOGÍA
P R E S E N T A :
CÉSAR JURADO ALANIZ

DIRECTOR: DR. RIGOBERTO LEÓN SÁNCHEZ

REVISORA: DRA. KIRARRSET BARRERA GARCÍA

SINODALES:

DRA. ZURAYA MONROY NASR

DR. OSCAR ZAMORA ARÉVALO

MTRO. FRANCISCO PÉREZ COTA

Tesis apoyada por el proyecto: "Las teorías implícitas de los
profesores universitarios sobre la enseñanza y el aprendizaje"
DGAPA PAPIIT IN306709

México, D.F.

2012





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

*Publicamos para no pasarnos la vida corrigiendo los
borradores.*

Alfonso Reyes

ÍNDICE:

AGRADECIMIENTOS	p. 9
INTRODUCCIÓN: Psicología y epistemología, la continuidad insalvable	p. 15
CAPÍTULO 1: Lenguaje, ciencia y psicología	p. 21
1.1. La filosofía y el estudio del comportamiento: una revisión a las relaciones entre lenguaje, epistemología y psicología	p. 22
1.2. El desarrollo de las psicologías y el desarrollo de la psicología: más allá del avance el áreas específicas de investigación	p. 30
1.3. Transgredir las fronteras: desdibujando la división entre ciencia y filosofía	p. 33

PRIMERA PARTE:

LOS PROCESOS COGNITIVOS COMO FENÓMENOS COMPUTACIONALES

CAPÍTULO 2: Los dos argumentos a favor del modelo del lenguaje: la mente, entre lo computacional y lo social	p. 43
2.1. Dos argumentos a favor del modelo del lenguaje	p. 44
2.1.1. El Argumento Intuitivo (AI) sobre la naturaleza lingüística de los procesos psicológicos superiores.	p. 45
2.1.2. El Argumento Sofisticado (AS) sobre la naturaleza lingüística de la cognición.	p. 51
2.2. Las dos metáforas lingüísticas de la mente: el libro y la computadora.	p. 60
CAPÍTULO 3: El AS en la ciencia cognitiva contemporánea: el estudio de la cognición específica de dominio.	p. 65
3.1. Módulos mentales y teorías implícitas: la especificidad de dominio y sus implicaciones para la comprensión de la arquitectura mental	p. 66
3.1.1. La organización modular de la facultad de lenguaje	p. 67
3.1.2. La arquitectura modular de los sistemas mentales periféricos	p. 70
3.1.3. Desarrollos actuales en modularidad mental y dominio especificidad	p. 72
CAPÍTULO 4: Modelos matemáticos en el estudio de los procesos cognitivos: el AS y el análisis experimental de la conducta	p. 85

4.1. El contraataque empirista: el análisis experimental de la conducta	p. 87
4.2. Los modelos matemáticos, ¿son sustitutos de las descripciones verbales de los fenómenos psicológicos?	p. 92
4.3. Modelamiento matemático y algoritmos computacionales	p. 104

SEGUNDA PARTE:

RAZÓN Y LÓGICA: HACIA UNA TEORÍA (REALMENTE) NATURALISTA DEL RAZONAMIENTO HUMANO

CAPÍTULO 5: Preliminares para una teoría naturalista del conocimiento humano	p. 117
5.1. Razón y lógica.	p. 121
5.2. Razón, más allá de la lógica y la matemática	p. 123
5.2.1. Lo básico y no tan básico del aprendizaje asociativo	p. 128
5.1.2. Sistemas duales de razonamiento: evolución y cognición	p. 135
5.3. La naturaleza interactiva de los sistemas cognitivos	p. 138
5.4. Las múltiples caras de la racionalidad	p. 146
CAPÍTULO 6: ¿“Adictos” a la información? La racionalidad como fenómeno natural	p. 153
6.1. ¿Continuidad o discontinuidad cognitiva? Las caras del problema	p. 155

6.2. Cuando el más es menos o ¿no lo pienses demasiado!	p. 159
6.3. Percepción y acción: el inconsciente cognitivo.	p. 170
6.4. Hacia una nueva teoría naturalista del la racionalidad humana:rel razonamiento y la vida en sociedad	p. 188
CONCLUSIONES	p. 207
Referencias	p. 129

AGRADECIMIENTOS

We are accidents waiting to happen

There There, Radiohead

Siempre he tenido la impresión de que las secciones más leídas de las tesis de licenciatura son, precisamente, los agradecimientos y la bibliografía, y con justa razón: tan importante como el producto son aquellos que lo hicieron posible. Hoy no estaría escribiendo esto de no ser por las personas con las que he tenido la fortuna de compartir mi vida (agradecimientos) y aquellos de cuyas ideas me he valido para desarrollar mis propias concepciones de los fenómenos psicológicos (bibliografía). Este trabajo es de ellos y para ellos.

AL COMITÉ DICTAMINADOR DE ESTA TESIS. A todos, mi admiración profunda y mi gratitud por la paciencia e impulso que mostraron hacia mí a lo largo del (un tanto caótico) proceso de escritura de este trabajo.

En el primer lugar, todo mi reconocimiento y respeto al Dr. Rigoberto León Sánchez por ser un maestro en todo el sentido de la palabra. Sus enseñanzas (tanto en sus clases y seminarios como en la interacción cotidiana) son en gran medida la causa de que esta sea una tesis de psicología y no de literatura, biología o cualquier otra área. Gracias por todo su apoyo, confianza y paciencia, en verdad que no me alcanzan las palabras para expresar cuanto le debo como profesional. Gracias también a la Dra. Kirareset Barrera García por leer las múltiples “versiones preliminares” de este y otros trabajos. Tengo la impresión de que nadie se pudo formar una idea más clara de lo importante que esta tesis resulta para mí. Gracias por propiciar el desarrollo de mis ideas y por interceder por mí cuando los tiempos del mundo no se correspondían con los tiempos en mi cabeza.

Gracias a la Dra. Zuraya Monroy Nasr por la precisión de sus comentarios, por revisar minuciosamente el escrito y por todas sus observaciones y recomendaciones; me fueron de mucha utilidad. Toda mi admiración para el Mtro. Francisco Pérez Cota y el Dr. Óscar Zamora Arévalo; en verdad que me siento muy honrado de tener en mi comité a los profesores que más marcaron mis concepciones sobre la psicología. Muchas de las ideas que desarrollo en la tesis nacieron con las discusiones que tuvieron lugar en sus grandiosos cursos.

A MI FAMILIA. Todo mi amor, admiración y respeto para mis padres, Leonor y Manuel, por su apoyo incondicional, a pesar de lo extraños que puedan resultar mis intereses. Mil gracias por velar siempre por el bienestar mío y

de mis hermanos, por todos los sacrificios que han tenido que realizar para que yo haya podido estudiar una licenciatura. De ustedes he aprendido mil cosas, pero ninguna tan importante como que no se puede ser un buen profesional sin antes ser una buena persona.

Mil gracias y todo mi amor para mis hermanos, Jorge, Olivia, Guillermo y Francisco. En ustedes siempre he tenido a mis mejores amigos. Gracias por todas las risas en los momentos alegres y por todo su apoyo en los momentos difíciles. A pesar de que cada uno sea tan diferente de los otros, estoy convencido de que somos los mejores hermanos del mundo y no hay nada que me enorgullezca más en la vida. Mi cariño para mi “cuñis”, Teresa, y para mis sobrinos Ahtziri y Rodrigo; gracias por traer felicidad a la vida de mi hermano y, de paso, a la de toda la familia Jurado.

A LOS “ANDAMIOSOS”: Todos mi afecto, admiración y respeto al Dr. Juan Manuel Sánchez, así como a Fabiola Rodríguez, Marisol de Diego, Karina Hurtado y a todos los miembros de proyecto “Andamios Curriculares”. Mi agradecimiento por todas las enseñanzas (profesionales y personales) que me dejaron a lo largo del tiempo que tuve la dicha de colaborar con ustedes. Gracias por confiar en mí, aún para realizar tareas que implicaban interactuar con otras personas, cosa que ustedes saben que no se me da muy bien. Gracias por las charlas académicas pero sobre todo por las no académicas; en verdad que me permitieron enriquecer mucho mi manera de comprender la vida y la psicología.

A MIS “CUATES” Y COMPAÑEROS: Todo mi afecto a Elizabeth Flores, Elizabeth Jiménez, Montserrat Vanegas, Julián Solís y Luis Abel León por ser unos excelentes amigos; por los muchos momentos gratos y las pláticas amenas, por el apoyo que mostraron hacia mí en los momentos en que las cosas no me salían del todo bien; en verdad que los estimo mucho y me siento muy afortunado de que sean mis amigos. Gracias a Carolina García por ser mi confidente, por soportar todas mis “crisis existenciales” y por jalarme las orejas cada vez que la situación lo amerita (no lo dejes de hacer, por favor); eres una gran amiga Carito, en verdad que no tengo más que gratitud y admiración hacia ti.

Gracias a Esmeralda Fonseca, Luis Abel León, Aline Moreno (Ross), Eduardo Parra, José Luis Baroja, Jaime Cadena y Adriel Ruiz por compartir su inteligencia conmigo. Es mucho lo que esta tesis le debe a las intensas y estimulantes charlas teóricas de explanada y pasillo que compartimos. Es muy grato conocer personas críticas y apasionadas con sus temas; no puedo decir más que “siempre es un placer estar en desacuerdo con ustedes”.

Mi gratitud a Jean Molina por ser un excelente compañero. Es poco el tiempo que tengo de conocerte pero en verdad que eres una gran persona. Gracias por siempre estar dispuesto a apoyarme en mis enredos académico-administrativos y por las charlas y los chistes filosóficos.

Mi admiración y respeto para Dalia Cisneros. Gracias por confiar en mí y por impulsarme en todos los aspectos durante el tiempo que tuve la fortuna de compartir contigo. Mi agradecimiento por escucharme con paciencia cuando me surgía una “gran idea psicológica” y por la elegancia y

sencillez con la que casi siempre me mostrabas que después de todo no era tan buena; hablar contigo siempre fue un gran estímulo para el desarrollo de ésta, la “mítica tesis”. Gracias, sobre todo, por recordarme que más allá de la psicología y la filosofía también hay un mundo (y que, a veces, hasta vale la pena darse una vuelta por él) y por mostrarme, muchas veces sin proponértelo, que al final del día lo más importante de la vida es precisamente aquello que se resiste a ser explicado, aquello que sólo se siente.

A MIS PROFESORES: Mil gracias a todos los profesores que realmente estuvieron dispuestos a compartir sus conocimientos con nosotros, sus alumnos, a lo largo de la carrera. Muy en especial a la Dra. Alejandra Ruiz Contreras, a la Dra. Irma Yolanda del Río, al Dr. Hugo Sánchez Castillo, a la Lic. Blanca Reguero Reza, al Dr. Pablo Fernández Christlieb y al Dr. Gustavo Ortiz Millán. Más allá de qué tan vinculados pudieron haber estado sus cursos con mis intereses, siempre admiraré la pasión con la que imparten sus temas y su compromiso con la docencia. Todos ustedes dejaron una huella muy profunda en mi formación, por ello, toda mi gratitud.

A LA UNIVERSIDAD NACIONAL: Gracias por el apoyo recibido a través de proyecto “Las teorías implícitas de los profesores universitarios sobre la enseñanza y el aprendizaje” **DGAPA PAPIIT IN306709**; la elaboración de esta tesis se vio muy favorecida con él. Sobre todo, gracias por regalarme los años más fascinantes de mi vida hasta ahora, siempre rodeado por un ambiente académico y cultural tremendamente estimulante.

INTRODUCCIÓN

PSICOLOGÍA Y EPISTEMOLOGÍA, LA CONTINUIDAD INSALVABLE

The limits of my language mean the limits of my world

Ludwig Wittgenstein

Esta tesis pretende dar respuesta a dos preguntas fundamentales: (1) ¿cuál es la relación entre nuestras teorías del lenguaje y nuestras teorías de la ciencia?, y (2) ¿cuál es la que prima entre nuestras teorías de la ciencia y nuestras teorías de la mente? Una consideración minuciosa del desarrollo de estas cuestiones nos revela que, contrario a lo que se podría esperar, la especificación de la naturaleza del lenguaje antecede a la caracterización del conocimiento científico y, aún más, a la descripción de la naturaleza de los fenómenos mentales.

A primera vista esto podría pasar por curiosidad histórica sin relevancia; no obstante, sostengo que un análisis más detallado revela que este hecho terminó por configurar la forma en que comprendemos los procesos psicológicos en la ciencia cognitiva contemporánea. El argumento que desarrollo a lo largo de este trabajo es que nuestras ideas, tanto de la ciencia

como de la mente, están marcadas por la filosofía formalista del lenguaje. De este modo, el lenguaje es visto como un puente entre el mundo y nuestra realidad psicológica, y el *conocimiento* que a partir de él podemos formar (la *ciencia*, particularmente hablando) sólo es posible en la medida en que nuestra mente puede entrar en contacto con la realidad por medio del lenguaje. Así pues, se defiende (muchas veces de forma implícita) la existencia de cierto *isomorfismo* en la *estructura lógica* de cada uno de los niveles de la empresa epistemológica: sólo podemos conocer el mundo en virtud de que su *estructura* guarda una fuerte correspondencia con las *estructuras* (¿lógicas?) que supuestamente constituyen nuestro lenguaje, nuestra cognición y nuestro conocimiento.

Del mismo modo que se defiende que el carácter significativo del lenguaje recae en ser un sistema de reglas para la combinación de símbolos y para la transformación de secuencias de símbolos en otras secuencias significativas, se admite que la validez del conocimiento científico depende de su capacidad de vincular lógicamente proposiciones sobre la estructura del mundo. En un momento posterior, con la llamada “revolución cognitiva”, los procesos mentales también pasarían a ser entendidos en términos lógicos, es decir, como fenómenos computacionales consistentes en la aplicación de reglas sobre representaciones discretas.

La tesis está estructurada en dos partes que engloban seis capítulos. El capítulo uno, de carácter mayormente histórico e introductorio, revisa la forma en que la filosofía positivista de la ciencia impactó en la manera en que en psicología se comprende tanto la naturaleza de su objeto de estudio

(los procesos mentales) como el carácter mismo de la empresa científica. Sostengo que el énfasis de los empiristas lógicos sobre los datos sensibles y la experimentación, así como su menosprecio por la teorización, provocaron que se estableciera una frontera tajante entre los esfuerzos de filósofos y científicos, privando (las más de las veces) a unos y a otros de la posibilidad de aprovechar los desarrollos e ideas surgidos de tradiciones intelectuales más allá de la propia. Argumento que la única forma en que puede avanzar nuestra comprensión de la cognición (y de la naturaleza en general) es a partir de la abolición de estas fronteras artificiales entre áreas del saber.

La primera parte está constituida por los capítulos dos, tres y cuatro. Aquí muestro la forma en que el “modelo del lenguaje” impactó la investigación empírica en ciencia cognitiva, particularmente en lo que tiene que ver con la conceptualización de los fenómenos mentales en términos computacionales. El capítulo dos delimita aquello a lo que me estoy refiriendo cuando hablo de “modelo del lenguaje”, a decir, la postura que defiende que tanto la ciencia como la mente son, en esencia, sistemas simbólicos especificables a partir de las mismas categorías que los lenguajes formales.

Los capítulos tres y cuatro constituyen una revisión más o menos detallada de la forma en que el “modelo del lenguaje” se ve reflejado en las tradiciones de investigación más prominentes dentro de la psicología cognitiva: la teoría computacional de la mente (capítulo 3) y el análisis experimental de la conducta (capítulo 4). En el capítulo cuatro, adicionalmente, argumento que es necesario diferenciar entre las expresiones formales

que pretenden representar *algoritmos cognitivos* y aquellas que intentan establecerse como descripciones de *procesos cognitivos*. Sostengo que el supuesto de que la ciencia y la mente son, por igual, formalmente especificables ha llevado a que los investigadores muchas veces pierdan ésta muy básica, pero también muy importante, distinción.

En la segunda parte, compuesta por los capítulos cinco y seis, abordo el problema del impacto de la lógica sobre nuestras teorías de la racionalidad. El capítulo cinco presenta de forma por demás escueta algunas de las líneas de investigación que, desde las propias ciencias cognitivas, nos exigen que modifiquemos la forma en que tradicionalmente hemos entendido la relación entre lenguaje, cognición y comportamiento, además de avanzar algunas ideas sobre la mejor manera en que se debería entender la relación entre razonamiento y conducta. En el capítulo seis, finalmente, argumentó que, dado que es insostenible la idea de que todas nuestras acciones son el resultado de procesos racionales, nuestra capacidad de representarnos explícitamente las relaciones de los eventos en el mundo debería jugar un papel distinto en nuestra vida mental del que tradicionalmente le ha sido asignado en las teorías psicológicas clásicas.

En suma, en esta tesis sostengo que es necesario que nuestras ideas sobre la cognición dejen de depender de la forma en que comprendemos la naturaleza de la ciencia. Pero también argumento que la única forma en que podemos apreciar con plenitud la fascinante naturaleza del saber humano (de cual la ciencia es un tipo particularmente interesante) es a partir de una comprensión profunda de nuestros procesos cognitivos y de las interaccio-

nes sociales que ellos posibilitan. Finalmente, una teoría psicológica que sea omisa al problema del conocimiento siempre será, por decir lo menos, una teoría psicológica incompleta. Dicho con otras palabras, la única forma en que podemos iluminar la relación entre cognición y conocimiento es dejar de considerar a una y otro como sistemas simbólicos explicables en términos puramente formales.

En este trabajo perfiló los elementos básicos que considero necesarios para una teoría naturalista de la racionalidad que dé luz tanto de su dimensión social, presente en instituciones como la ciencia, como de su dimensión individual, necesaria para comprender la relación entre mundo, cognición y comportamiento. Aunque es innegable que numerosos autores han realizado importantes aportaciones en esta dirección, pienso que en los últimos años la investigación en ciencias cognitivas está abriendo nuevos y fascinantes caminos que merecen ser explorados.

Mi énfasis recae sobre la urgencia de disociar los mecanismos implicados en la racionalidad científica (de naturaleza eminentemente colectiva y distribuida) de aquellos que entran en juego cuando agentes individuales hacen frente al entorno inmediato (incluidos científicos y filósofos haciendo ciencia y/o filosofía), así como en la necesidad de clarificar la forma en que los procesos sociales de construcción del conocimiento pueden estar soportados en las limitadas capacidades cognitivas de los miembros de nuestra especie, pero sin reducirse a ellas.

Esta tarea de redefinición de nuestros criterios epistemológicos es relevante no sólo desde el punto de vista teórico, sino que tiene implicaciones

directas y profundas sobre la forma en que se lleva a cabo la investigación científica y en cómo se comunica el conocimiento surgido de ella tanto al interior como al exterior de los círculos académicos.

CAPÍTULO UNO

LENGUAJE, CIENCIA Y PSICOLOGÍA.

If science and philosophy refer to methods, then these methods work in concert and are anemic when performed in isolation. Data without theory is empty, and theory without data is blind.

Jesse J. Prinz

Como primer paso para comprender la relación entre lenguaje, ciencia y cognición, resulta necesario revisar el contexto intelectual que propició que estos tres fenómenos pasaran a estar tan estrechamente relacionados que resulta prácticamente imposible referirse a uno sin considerar a los otros dos. Agotar esta cuestión podría remitirnos al origen mismo de los registros de nuestro conocimiento, no obstante, considero que uno de los momentos claves de esta historia —al menos aquel que está más estrechamente vinculado con la forma en que este problema es comprendido en la actualidad— es el nacimiento mismo de la filosofía de la ciencia con las consideraciones desarrolladas por los miembros del círculo de Viena a principios del siglo XX.

Este primer capítulo tiene como objetivo reconstruir ese contexto intelectual para, a partir de ahí, apreciar la forma en que nuestras ideas sobre la

cognición están marcadas, *desde su mismo origen*, por nuestras ideas sobre el lenguaje y la ciencia. Como será claro a lo largo del desarrollo de mis reflexiones, la influencia de la epistemología sobre el estudio de la mente es mucho más compleja y multidimensional de lo que podría presentársenos en un primer momento. Simplificando el problema con fines meramente expositivos, podría decirse que el empirismo lógico impacta por igual tanto nuestras concepciones de la naturaleza y alcance de la empresa científica como la caracterización a partir de la que partimos a la hora de encarar el estudio de los fenómenos mentales.

Como un segundo momento en el desarrollo de este capítulo introductorio, enfatizo la importancia de que trascendamos la burda distinción entre saberes filosóficos y saberes científicos. Como muy bien lo hace ver Jesse Prinz (2008), la ciencia y la filosofía no deben ser vistas como sistemas antagónicos de conocimiento, sino, por el contrario, como métodos complementarios que, a partir de su puesta en marcha como directrices de nuestras prácticas epistémicas, nos permitan ganar una mayor comprensión de la realidad que nos circunda y de nuestro lugar dentro de ella.

1.1. LA FILOSOFÍA Y EL ESTUDIO DEL COMPORTAMIENTO: UNA REVISIÓN A LAS RELACIONES ENTRE LENGUAJE, EPISTEMOLOGÍA Y PSICOLOGÍA

Si uno tuviera que encontrar una palabra para caracterizar el desarrollo intelectual que tuvo lugar en occidente a lo largo del siglo pasado, una de las más fuertes candidatas habría de ser la palabra “lenguaje”. Como nunca

antes, lo lingüístico se coloca al centro de las discusiones teóricas, y esto con independencia del campo de conocimiento en que éstas se desarrollen. Desde el código genético en biología y hasta el auge de disciplinas como la informática y la computación, el siglo XX (en que los alcances de nuestra comprensión sobre el mundo se extendieron a una velocidad y un número de áreas sin precedente en nuestra historia como especie) es simplemente impensable sin la particular distribución de ocho letras omnipresentes en los registros de nuestro conocimiento: L-E-N-G-U-A-J-E.

Si bien es innegable que el lenguaje ha jugado desde siempre un rol muy importante en nuestras caracterizaciones de la realidad (véase, por ejemplo, el capítulo 8 en Olson, 1994/1998 para una excelente revisión del desarrollo de la concepto de ciencia como la lectura del “libro de la naturaleza” a lo largo del siglo XVII), a partir del llamado “giro lingüístico” éste pasó a convertirse en *modelo*, *medio* y *medida* de nuestro conocimiento sobre el mundo.

Esto es, el lenguaje se convierte en *modelo* en el grado en que se pretende que buena parte de los fenómenos naturales pueden ser descritos a partir de las mismas categorías empleadas en la caracterización de los lenguajes formales. Pasa a ser *medio* tanto en los casos en que se acepta que la realidad en su conjunto, o la ciencia como estudio de la primera, se organizan de manera lingüística. Y, por último, el lenguaje se constituye como *medida* cuando la *validez* de nuestro conocimiento descansa sobre “su correcta estructuración” evaluada a partir del supuesto “metalenguaje neutral” que en algún tiempo representó la lógica de primer orden.

Si bien este *giro lingüístico* es generalizado, su influencia es aún más dramática en las disciplinas orientadas al estudio del comportamiento y la cognición, consideraciones que abordaré a profundidad a lo largo de este trabajo.

Aunque las aportaciones de este cambio al progreso de la ciencia son invaluableles, no son pocas las veces en que los desarrollos teóricos que en algún punto fueron el motor que impulsó el florecimiento de una disciplina terminan por convertirse en los lastres que obstaculizan la continuidad de su desarrollo. De este modo, es de vital importancia determinar hasta qué punto el paradigma del lenguaje puede continuar soportando el avance de nuestra comprensión de los fenómenos cognitivos.

Entiendo que no hay mayor expresión de este énfasis lingüístico que la filosofía desarrollada por los miembros del Círculo de Viena durante las primeras décadas del siglo XX. En la actualidad resulta difícil suponer que haya alguien que sostenga las tesis positivistas, al menos como fueron planteadas en un primer momento. Mucha tinta ha corrido en torno a la eficacia y conveniencia de una postura epistemológica como la defendida por el positivismo lógico y este no es lugar para abundar en estas polémicas. Si aquí se consideran las discusiones propias de la filosofía analítica es sólo en tanto son relevantes para entender su influencia sobre la forma en que se realiza la investigación en las disciplinas interesadas en comprender la forma en que los organismos interactúan, representan y dan respuestas eficientes a las demandas impuestas por el ambiente. En última instancia, *de lo que se trata es de determinar hasta qué punto el modelo del lenguaje,*

con sus supuestos y exigencias, sigue siendo fructífero y adecuado para caracterizar los fenómenos psicológicos.

Más aún, es importante tener presente que el giro lingüístico es un acontecimiento que se extiende mucho más allá de una disciplina, una teoría o un momento histórico particular y constituye uno de esos fenómenos culturales que abarca a todo el pensamiento: *el giro lingüístico es, ante todo, una manera de mirar el mundo por medio del lenguaje* (como opuesto a ver el lugar del lenguaje entre los fenómenos del mundo). La filosofía y las ciencias contemporáneas no son menos lingüísticas que la postura de la cual se desprenden y el desmoronamiento del programa positivista no hace más que complejizar las interacciones entre lenguaje, naturaleza y ciencia, haciendo más laboriosa la tarea de determinar las formas en que el primero media en la aproximación a la segunda y la conceptualización de la tercera. Un paso crucial para el progreso del estudio de la cognición y el comportamiento consiste en hacer explícitas las múltiples influencias del Lenguaje en el estudio y la conceptualización de los fenómenos psicológicos.

El caso de la psicología es bastante ilustrativo de las interrelaciones entre teorización epistemológica y ciencias empíricas particulares. Desde el comienzo mismo del programa positivista, nuestra disciplina se constituyó como un blanco natural tanto de sus críticas como de sus reformulaciones. Por un lado, las teorías psicológicas dominantes a principios del siglo XX representaban el tipo de explicaciones que, de acuerdo con la epistemología basada en el análisis lógico del lenguaje, tenía que ser abandonadas por carecer de sentido (Carnap, 1932/2002; Hempel, 1931/1980).

Por otra parte, la psicología constituye la bandera con la cual el positivismo lógico se lanza contra las posturas que defienden una frontera tajante entre los objetos de estudio, métodos y principios propios de las ciencias naturales (*Naturwissenschaften*) y aquellos pertenecientes al dominio de las llamadas “ciencias del espíritu” (*Geisteswissenschaften*) (Dilthey, 1890/1980). Como comenta Edward Schouten Robinson (1939):

En general, nosotros tendemos a creer que una disciplina apenas merece ser considerada como una “ciencia” al menos que esté basada sobre ciertos procedimientos empíricos que agrupamos bajo la etiqueta de “método científico”. Sin embargo, muchos escritores alemanes, siguiendo la postura propuesta por Wilhelm Dilthey hace cincuenta años, han tendido a distinguir dos tipos de *Wissenschaften* [ciencia] — *Naturwissenschaften* y *Geisteswissenschaften*— asumiendo que éstas difieren de manera radical tanto en su objeto de estudio como en sus procedimientos y que los métodos estrictamente empíricos son más apropiados para las primeras que para las segundas. Es contra esta división que los defensores del movimiento de la Unidad de la Ciencia han dirigido sus polémicas más vigorosas [...] (p. 129, corchetes añadidos).*

De esta manera, toda disciplina que no se ciña a los métodos “estrictamente empíricos” de la *Naturwissenschaften* no puede ser considerada como una ciencia. Además, se defiende que la existencia de disciplinas científicas particulares sólo es justificable como un abordaje provisional que facilite

* *Nota sobre las citas:* La traducción de las citas corrió a mi cuenta en los casos en que el idioma de la fuente, consignado en el apartado de referencias, sea distinto del español (al menos que se indique lo contrario). El formato de la traducción intenta respetar al máximo el énfasis de texto original (uso de cursivas, negritas, etc.); en caso contrario, se especifican los cambios realizados.

la aproximación al estudio de la naturaleza. En la medida en que las ciencias individuales se vayan desarrollando, éstas tenderán más y más hacia la integración hasta que por fin se pueda hablar *una ciencia* unificada que dé cuenta de la totalidad de los fenómenos naturales a partir de un conjunto de leyes unitarias (Neurath, 1937; Robinson, 1939).

El programa positivista va más allá al postular un principio que habría de regir su consecución: *el fisicalismo*. En un sentido estrecho, la unidad de las leyes que constituirían la ciencia unificada ya es una noción fisicalista: parte del supuesto de que la física es el modelo al que han de ajustarse el resto de las disciplinas de modo tal que, por ejemplo, las leyes que explicarían los fenómenos biológicos habrían de derivarse *lógicamente* de las leyes físicas que, a su vez, estarían soportadas en datos sensoriales. Por otro lado, se puede aplicar el término fisicalismo a otros dos aspectos de la empresa científica, y es esta aplicación la que constituye el supuesto camino que llevaría a la ciencia unificada: el fisicalismo puede suponerse como el eje rector tanto en el *método* como en el *lenguaje* de la ciencia. De este modo, se entiende que el método de la ciencia es uno y el mismo para todas las disciplinas: la aproximación empírica a los fenómenos de la naturaleza. En lo que tiene que ver con el lenguaje, éste también habría de ser unitario a la ciencia: el lenguaje extensional propio de la causalidad física (Carnap, 1932/2002; Neurath, 1937; Robinson, 1939). Como comenta Rudolf Carnap (1932/2002):

Si el lenguaje físico, sobre la base de su universalidad, fuera adoptado como el sistema de lenguaje (system language) de la ciencia, toda ésta pasaría a ser física. La metafísica habría de ser descartada por carecer de significado. Los distintos dominios científicos pasarían a ser parte de la ciencia unificada. En el modo material de hablar: habría, básicamente, sólo un tipo de objeto —ocurrencias físicas [...] (p. 39).

El impacto que estas consideraciones tuvieron sobre el desarrollo de la psicología fue dramático. Si las disciplinas interesadas en el estudio de los fenómenos sociales aspiraban a conservar su estatus de “ciencia” tendrían que ajustarse a los criterios establecidos por este fisicalismo. Casi al mismo tiempo que se desarrollaban estas discusiones, empezaron a surgir, desde la propia filosofía, propuestas concretas encaminadas a “liberar” a la psicología de la carga “metafísica” de sus teorías y dotarla de las herramientas y el rigor propios de la ciencia natural (véase, por ejemplo, Carnap, 1932/2002; Hempel, 1931/1980; Ryle, 1949/2005). En este sentido, Carl Hempel (1931/1980) afirma que:

Todos los enunciados psicológicos con significado, es decir, que sean en principio verificables, son traducibles a enunciados que no involucren conceptos psicológicos, sino sólo los conceptos de la física. Los enunciados de la psicología son, en consecuencia, enunciados fisicalistas. La psicología es una parte integral de la física (p. 18).**

** Traducción sin publicar de Gustavo Ortiz-Millán, Instituto de Investigaciones Filosóficas, UNAM, de la versión en inglés que apareció en Block, N. (1980) (ed.), *Readings in Philosophy of Psychology*. Cambridge, Mass.: Harvard University Press.

O, de acuerdo con Rudolf Carnap (1932/2002):

Las diferentes ciencias han alcanzado hasta ahora muy diferentes etapas en su proceso de descontaminación metafísica. Gracias principalmente a los esfuerzos de Mach, Poincaré y Einstein, la física está, en general, prácticamente libre en este sentido. En la psicología, por otra parte, el proceso de arribar a una ciencia libre de metafísica apenas está comenzado. La diferencia entre estas dos ciencias es vista de forma más clara en las diferentes actitudes tomadas por los expertos en cada una de estas disciplinas hacia la posición que nosotros rechazamos como metafísica y carente de significado (p. 44).

Así pues, para que los enunciados psicológicos tengan sentido han de estar formulados a través del lenguaje puramente extensional (“universal e intersubjetivo”) de la física. En lo específico, el lenguaje mentalistas propio de los enunciados psicológicos habría de poder ser *reducido* (traducidos sin pérdida de contenido) al lenguaje de las manifestaciones tangibles y empíricamente abordables de la conducta manifiesta y los estímulos ambientales (Carnap, 1932/2002; Hempel, 1931/1980; Ryle, 1949/2005).

Si el tono de estos argumentos resulta familiar no habría de ser sorprendente. En gran medida este “conductismo lógico” sirvió como fundamento conceptual (e ideológico) del conductismo psicológico defendido por autores como John B. Watson (1930/1970) y Burrhus Frederic Skinner (1938/1975), que es, a su vez, la materialización del abordaje empírico propuesto por el primero.

Si ya desde siempre los caminos de la epistemología y la psicología habrían estado ligados de manera muy estrecha, justo sería decir que a partir del giro lingüístico estos pasaron a ser hasta tal punto interdependientes que, en términos prácticos, todo desarrollo teórico en psicología halla su fundamento filosófico en algún punto de la discusión sobre el lenguaje y/o el conocimiento científico.

1.2. EL DESARROLLO DE LAS PSICOLOGÍAS Y EL DESARROLLO DE LA PSICOLOGÍA: MÁS ALLÁ DEL AVANCE EL ÁREAS ESPECÍFICAS DE INVESTIGACIÓN

Lo dicho hasta ahora podría dar la idea de que defiende una perspectiva puramente externalista de la evolución del saber psicológico. No obstante, esta es una apreciación muy simplista de lo que aquí he venido argumentando. Si bien joven, a lo largo de su historia nuestra disciplina ha acumulado un cuerpo impresionante de hallazgos empíricos y desarrollado modelos teórico-explicativos con un impresionante poder predictivo sobre la base de este conocimiento. La psicología puede, como hasta ahora, seguir enriqueciendo y desarrollando sus modelos teóricos y, aún más, proponiendo unos nuevos y más poderosos *en aspectos particulares del extenso dominio de los fenómenos psicológicos*. La crítica que aquí desarrollo no tiene nada que ver con lo “fructíferas” (en un sentido estrecho de la palabra) que pueden resultar las propuestas teóricas en psicología, si no que va dirigida a la *falta de integración existente entre éstas*. De modo crucial, no existen criterios claros a partir de los que se pueda evaluar la pertinencia

de algún determinado modelo psicológico o, lo que es más, el papel de un determinado hallazgo empírico en la supuesta *explicación general* de algún fenómeno.

La investigación psicológica está plagada de sofisticados modelos que pueden llegar a ser muy robustos para explicar el comportamiento de los organismos *ante un determinado paradigma experimental* pero que pierden toda su elegancia cuando de lo que se trata es de echar mano de ellos para dar cuenta de la conducta que tiene lugar más allá de los estrechos límites del laboratorio. Buena parte de la investigación de los procesos cognitivos se va en acentuar esta atomización de los fenómenos mentales sin que exista un esfuerzo correspondiente para integrar los diversos modelos teóricos en una explicación coherente y sensible a la realidad ecológica dentro de la que tiene lugar la cognición y el comportamiento.

A pesar de ser herederas directas de la misma tradición intelectual que defienden el rigor y el ajuste de las explicaciones a los datos empíricos, nuestras propuestas teóricas siguen desarrollos teóricos y conceptuales mayormente independientes, *aún teniendo intereses muy próximos, y a veces hasta comunes, en el espectro de los fenómenos psicológicos*. Este hecho se hace aún más evidente cuando consideramos la posibilidad de integración (o, por decir lo menos, genuina discusión) entre perspectivas teóricas surgidas de tradiciones diferentes pero que comparten el interés en un mismo objeto de estudio (véase, por ejemplo, los comentarios de Wasserman y Zentall, 2006 con respecto a las diferencias entre las áreas de cognición comparada y etología cognitiva).

Es posible que haya muchos que vean estas consideraciones como exageradas e imprecisa en muchos aspectos. Por ejemplo, se puede argumentar que muchos de los teóricos más prominentes en un campo también han contribuido de forma invaluable al desarrollo de otras áreas. Pero lo que está aquí a discusión no es la genialidad de individuos particulares, sino el grado en que podemos decir que estamos —como disciplina— cerca de una *explicación general* de un fenómeno como, digamos, el aprendizaje o, menos aún, si tenemos alguna idea de qué *características* habría de tener una explicación de esta naturaleza. De manera concreta, esto es a lo que me refería cuando hablaba de una falta de criterios para evaluar los desarrollos y las teorías: es difícil hallarles un lugar cuando se ve más allá de los estrechos límites del modelo que les dio origen.

Buena parte de la investigación psicológica es en la actualidad dependiente del diseño experimental o la tarea a partir de la que se lleva a cabo, y esto es a la vez causa y consecuencia de su falta de validez ecológica. Un diseño experimental puede constituirse como una realidad empírica hasta un punto en que los investigadores estén simplemente impedidos para dar cuenta de los datos empíricos más allá del modelo mismo, incapaces de considerar la posibilidad de explicaciones alternativas fuera de éste.

Podría ser del todo legítimo aspirar a esclarecer hasta el último particular del modo en que un organismo se comporta *en la situación concreta impuesta por un determinado paradigma experimental*. Sin embargo, no parece que sea esto lo que realmente motiva la investigación: en psicología, el objetivo es establecer las *leyes generales* que gobiernan la emisión de

la conducta. Cabría preguntarse, ¿a qué nivel de generalidad pueden aspirar nuestras explicaciones si en un grado importante están subordinadas al diseño experimental con la que nos aproximamos a los fenómenos y no de los fenómenos mismos?

La salud de la psicología como disciplina depende de mucho más que de la salud de sus programas de investigación particulares. Considero que el único camino que puede llevarnos a un genuino desarrollo de nuestra comprensión de los fenómenos cognitivos es una reflexión profunda en torno a nuestras concepciones sobre el conocimiento científico, el papel que la psicología juega dentro de él y la naturaleza de nuestros propios procesos cognitivos (que en última instancia nos permiten hacernos todas estas preguntas), seguida de la discusión frontal, permanente y verdaderamente informada de nuestras propuestas explicativas.

1.3. TRANSGREDIR LAS FRONTERAS: DESDIBUJANDO LA DIVISIÓN ENTRE CIENCIA Y FILOSOFÍA

¿Se puede aislar a los fenómenos psicológicos del resto de los fenómenos naturales? ¿Cuál debe ser la relación entre la psicología y el resto de las disciplinas científicas? ¿Cómo se relaciona la ciencia empírica con las discusiones conceptuales de la filosofía? Todas estas preguntas están vinculadas con el problema de comprender la naturaleza de conocimiento científico, en lo particular, y del conocimiento humano, en lo general, cuestiones que son de central importancia en el desarrollo de este trabajo.

La filosofía de la ciencia moderna ve la luz con el positivismo lógico desarrollado por los miembros del Círculo de Viena (Moulines, 2011). De acuerdo con estos autores, el conocimiento científico es el único género posible de genuino conocimiento; es decir, no hay más saber que aquel que es empíricamente verificable y/o deriva lógicamente de proposiciones empíricamente verificables. Las teorías científicas son entendidas como sistemas de proposiciones lógicamente articuladas, desarrolladas por agentes totalmente racionales (los científicos) al margen de cualquier otra práctica cultural.

No deja de resultar paradójico que la “descontaminación metafísica” del conocimiento defendida por autores como Rudolf Carnap (1932/2002) o Edward Robinson (1939) conlleve el rechazo de las preocupaciones, pretensiones y métodos que alimentaron durante siglos el desarrollo del conocimiento filosófico y, en consecuencia, el abandono de casi la totalidad de éste. Para los empiristas lógicos, si la filosofía habría de jugar papel alguno en el progreso de nuestra comprensión del mundo, éste estaría restringido al análisis lógico de las teorías científicas. Dicho de otro modo, el positivismo es la filosofía de la renuncia a la filosofía.

En su embelesamiento científicista, la epistemología contemporánea no se ha dado cuenta que, más que contribuir al desarrollo de nuestra comprensión del mundo, ha ayudado a establecer una frontera tajante entre los esfuerzos de científicos y filósofos, privando así a las disciplinas experimentales de la posibilidad de aprovechar las poderosas herramientas de análisis conceptual desarrolladas por la filosofía, y a los filósofos, a su vez,

de un suelo empírico firme sobre el cual edificar sus teorías. Y es que si bien los filósofos muestran gran interés en la ciencia, algunas veces parece que poco o nada les importa lo que hagan los científicos. De este modo, las discusiones filosóficas corren el peligro de convertirse en genuinas charlas bizantinas que sólo en ocasiones repercuten sobre otras formas de explorar la realidad.

En este punto, ha de ser claro que el impacto de la filosofía positivista va mucho más allá de un peculiar (e idealizada) caracterización de la ciencia; el empirismo lógico, conlleva una serie de compromisos ontológicos tremendamente fuertes sobre la naturaleza y la realidad, pero también (y sobretodo) a cerca del conocimiento que sobre ellas podemos formar. En todo este esquema, permea la idea de que todos los niveles (la realidad, el pensamiento, el lenguaje y la ciencia) son *estructuras*, es decir, conjuntos de elementos relacionados a partir de un número finito de reglas, y que, en tanto “estructuras”, son sintácticamente especificables.

Como comentan Brun y Kuenzle (2008): “En el orden tradicional [del análisis epistemológico], las evaluaciones epistémicas de proposiciones, sentencias o estados mentales fueron realizadas primero y entonces los agentes, actos y procesos epistémicos fueron explicados en términos de estos análisis” (p. 6). Es decir, no es sólo que nuestros criterios epistémicos resulten ajenos a las condiciones reales del desarrollo de la ciencia, sino también que las ideas que sostenemos sobre los procesos de construcción del saber, las prácticas que intervienen en éstos y los agentes que las llevan a cabo terminaron por asimilar como propias las (idealizadas) cualidades

que se consideran características de la ciencia. Desde el Giro lingüístico “lenguaje”, “ciencia” y “mente” han pasado a ser tres formas de nombrar un mismo fenómeno: sintaxis.

Aunque los empiristas lógicos creían que su trabajo podría contribuir a la construcción de una *ciencia unitaria*, es decir, a una paulatina integración de los conceptos y métodos de las disciplinas empíricas particulares hacia un sistema teórico que nos permitiera dar cuenta de la totalidad de los fenómenos naturales a partir de unas cuantas leyes generales, lo cierto es que su énfasis sobre la experimentación y el análisis de datos sensibles, así como su menosprecio al papel de las discusiones conceptuales dentro de la construcción del conocimiento científico (en suma, su *verificacionismo*), tuvo el efecto opuesto de aislar a las tradiciones teóricas en torno a procedimientos experimentales particulares. Revisar la forma en que este hecho se refleja en el desarrollo de la psicología fue el objetivo de la sección anterior.

Resulta transparente que la mejor forma para superar este aislamiento es por medio de la colaboración abierta entre profesionales de diferentes disciplinas en torno a problemas que engloben sus intereses y que aprovechen sus habilidades. De este modo, los problemas psicológicos son indisolubles de las consideraciones propias de la biología evolutiva o de la neurofisiología y neuroquímica del cerebro. No obstante, persiste la ambigüedad con respecto a la relación que debe primar entre las discusiones filosóficas y la labor científica. En este trabajo recupero la propuesta que recientemente han avanzado Knobe y Nichols (2008, y los trabajos ahí

compilados) sobre la necesidad de diluir la distinción entre ciencia y filosofía. Más que sistemas aislados de conocimiento, ciencia y filosofía son dos sistemas interrelacionados *implicados* en la adquisición de conocimiento.

La posibilidad de un abordaje empírico y/o experimental a las cuestiones filosóficas no puede ser vista más que con buenos ojos siempre y cuando se tenga presente que la cuestión no está en sustituir los métodos tradicionales de la filosofía por los propios de la ciencia. Como comenta Jesse Prinz (2008):

Necesitamos gente que sea recompensada por recabar datos y gente que lo sea por reflexionar, a veces de forma muy especulativa, sobre lo que todos ellos significan [...] El mundo académico no debería ser visto como una colección de feudos intelectuales aislados, sino como una constelación de métodos y tradiciones de investigación que colectivamente pueden contribuir a nuestra comprensión del mundo y nuestro lugar dentro de él (p. 207).

De esta forma, el verdadero potencial de la aproximación empírica a los problemas filosóficos recae precisamente en que suma diversidad, no sólo al poner sus propios supuestos sobre la mesa, sino también a aproximar corrientes de pensamiento que, aún teniendo intereses comunes, se han venido desarrollando de forma mayormente aislada desde que se optó por establecer una frontera tajante entre ciencia y filosofía (Prinz, 2008).

Estas consideraciones, que resultan pertinentes para cualquier dominio intelectual, adquieren un carácter de urgente cuando se piensa en la relación entre psicología cognitiva y epistemología. No es sólo de que los re-

sultados surgidos de la investigación empírica sobre los procesos humanos de razonamiento y de transmisión y conservación del conocimiento puedan favorecer el desarrollo de propuestas epistemológicas mucho más creíbles y robustas, sino también de hacer evidentes las múltiples influencias que la epistemología ha ejercido desde los comienzos de la psicología y las otras ciencias cognitivas; influencias que a la postre terminaron por convertirse en verdaderos dogmas que actualmente obstaculizan el desarrollo de lo que sabemos sobre los fenómenos cognitivos. No se trata únicamente de *naturalizar la epistemología*, sino también de someter los supuestos que subyacen al estudio empírico de la cognición a la reflexión y crítica conceptuales que tanto pueden abonar a nuestra comprensión científica del mundo. En otras palabras, nos encontramos en un contexto intelectual propicio para promover propuestas teóricas que aprovechen las herramientas desarrolladas por filósofos y científicos cognitivos y nos permitan ganar claridad sobre la forma en que el conocimiento aparece, se distribuye y se transforma hasta dar lugar a sistemas tan complejos como la ciencia moderna. Como la cita de Jesse Prinz (2008) hace evidente, los fenómenos naturales no entienden de divisiones académicas, ¿por qué entonces habría de ser diferente con nuestros esfuerzos por comprenderlos?

En este capítulo argumenté que tener al lenguaje como modelo para explicar la cognición y la conducta representa un problema para la comprensión de los fenómenos mentales. Mostré de forma esquemática que el desarrollo de las discusiones teóricas en ciencia cognitiva replicó a aquellas que tuvie-

ron lugar con varias décadas de antelación dentro de la filosofía del lenguaje y, particularmente, en la filosofía de la ciencia. Argumenté también que el único camino posible para encarar de forma productiva el reto de ofrecer verdaderas explicaciones naturalistas de los procesos cognitivos (unas que resulten progresivamente más esclarecedoras y fructíferas) es la confrontación abierta de nuestras ideas. No hay razones para pensar que la explicación de los procesos cognitivos pueda ser agotada a partir de los conceptos y métodos propios de un campo particular de conocimiento. Dicho de otro modo, el criterio último de convalidación de nuestras teorías científicas debe de estar marcado por la naturaleza de los fenómenos que se pretende explicar, y no hay razones (más allá de cierto realismo metafísico ingenuo) para suponer que la estructura del mundo se corresponda fielmente con la demarcación administrativa de las áreas del saber.

PRIMERA PARTE:

LOS PROCESOS COGNITIVOS COMO FENÓMENOS COMPUTACIONALES

CAPÍTULO DOS

LOS DOS ARGUMENTOS A FAVOR DEL MODELO DEL LENGUAJE: LA MENTE, ENTRE LO COMPUTACIONAL Y LO SOCIAL

The relation between thought and word is a living process; thought is born through words. A word devoid of thought is a dead thing. But thought that fails to realize itself in words also remains a “Stygian shadow”.

Lev Vygotsky

El actual panorama de la investigación psicológica no puede ser cabalmente apreciado sin hacer mención del contexto intelectual en que se gestaron sus problemas. Aunque es risible afirmar que cada teórico interesado en la comprensión de los procesos cognitivos ubique el origen de sus inquietudes en las reflexiones filosóficas sobre la naturaleza del lenguaje y la ciencia, sostengo que es más que claro que los *supuestos* —los cimientos mismos sobre los que nació y se viene desarrollando la ciencia cognitiva— son herederos directos de una tradición de pensamiento que hizo del lenguaje la medida del resto de los fenómenos naturales. El estancamiento de nuestra comprensión de los procesos de pensamiento no emerge tanto de las limitantes inherentes al modelo del lenguaje, sino de la *aceptación acrítica* de una serie de nociones sobre la naturaleza de la cognición y la mejor forma de explorarla empíricamente.

En este segundo capítulo desarrollo con más precisión a qué me estoy refiriendo cuando hablo de “modelo del lenguaje”. Como primer paso en esta dirección, encuentro necesario distinguir entre dos líneas principales de argumentación a favor del modelo del lenguaje, así como explorar la forma en que ambos tipos de argumentos suelen interactuar dentro de las justificaciones teóricas de los procesos psicológicos.

2.1. DOS ARGUMENTOS A FAVOR DEL MODELO DEL LENGUAJE.

Tal cual lo he venido manejando, podría parecer que al hablar de “modelo del lenguaje” estoy abarcando nociones demasiado heterogéneas como para ser englobadas en un concepto que sea productivamente abordable desde el punto de vista teórico. Aunado a esto, a primera vista, muchos estudiosos de la cognición podrían simplemente rechazar la posibilidad de que en su actividad cotidiana echen mano de categorías emanadas del estudio del lenguaje.

Es por esto que antes de exponer la forma en que el modelo del lenguaje se refleja en la teorización psicológica contemporánea considero necesario delimitar a qué me estoy refiriendo cuando sostengo que “es necesario dejar de caracterizar a la cognición como si fuera una entidad lingüística”. Así pues, encuentro preciso diferenciar entre dos líneas principales de argumentación que hay que tener presentes cuando se pretende esclarecer “qué tan lingüísticas” son nuestras caracterizaciones de la mente: el argumento intuitivo y el argumento sofisticado a favor del modelo del lenguaje.

2.1.1. EL ARGUMENTO INTUITIVO (AI) SOBRE LA NATURALEZA LINGÜÍSTICA DE LOS PROCESOS PSICOLÓGICOS SUPERIORES.

La primera línea de argumentación, que englobaré bajo la etiqueta de “Argumento Intuitivo” (AI), supone que la cognición (o, dicho de otro modo, “los procesos psicológicos superiores”, o simplemente “el pensamiento”) sólo pueden tener lugar a partir del dominio de un lenguaje natural. Desde esta perspectiva, la relación entre lenguaje y pensamiento es tan directa y estrecha que no parece tener sentido hablar del pensamiento sin hablar del lenguaje: el pensar es, esencialmente, una cuestión de representarse el mundo lingüísticamente.

Buena parte de la fuerza y atractivo de este tipo de argumentos radica, precisamente, en su plausibilidad de cara al sentido común: al pensar en la naturaleza de mi propio proceso de razonamiento, me resulta transparente y directamente accesible que mis pensamientos se estructuran como sentencias expresadas en mi lengua materna. Sin embargo, el alcance de esta idea trasciende ampliamente la apelación a la introspección. El supuesto de que la posesión de un lenguaje *posibilita* el pensamiento (o al menos la clase de pensamiento interesante desde el punto de vista de la psicología científica) está en la base de una gran diversidad de propuestas teóricas (filosóficas y psicológicas) que, sobre decirlo, suelen ser muy sofisticadas y sólo tienen de intuitivo este principio básico que las sustenta.

José Luis Bermúdez (2003) presenta como uno de los pilares de esta tradición argumentativa a las nociones derivadas de los escritos de Gottlob

Frege sobre relación entre el pensamiento y los lenguajes naturales. Para Frege, los pensamientos no son más que el *sentido* de las sentencias que los instancian: “Esos pensamientos son objetos abstractos que sirven no simplemente como objetos de actitudes proposicionales, tales como el conocer o el creer, sino también como los *sentidos* de sentencias” (Bermúdez, 2003, p. 14, cursivas en el original). Se entiende que los pensamientos adquieren su carácter *individual y estructurado* a través del acto lingüístico que les dio origen, siendo así susceptibles de entablar *relaciones lógicas* entre ellos. Esta propiedad de poder ser modelado por medio de la lógica proposicional se considera la esencia misma del proceso de pensar.

Estamos aquí frente al nacimiento de una de las tradiciones intelectuales más influyentes de la historia del saber occidental. Frege iguala la naturaleza de los lenguajes naturales con las categorías propias de los lenguajes formales y, a su vez, con la estructura total del proceso del pensamiento. No obstante, es importante tener presente que este “espíritu formalista” tiene repercusiones más profunda en las argumentaciones sofisticadas a favor del modelo del lenguaje, por lo que me reservaré la clarificación de sus implicaciones hasta el apartado siguiente.

Por el momento, me gustaría concentrarme en la presentación de una propuesta teórica mucho más familiar en el contexto de la discusión psicológica: las ideas del psicólogo soviético Lev Vygotsky sobre el desarrollo de los “procesos psicológicos superiores”. Tomo el proyecto vygotkiano como particularmente representativo de esta primera línea de argumentación dado que, considero, no hay exposición más clara y elegante de la

forma en la que la adquisición del lenguaje transforma la cognición que las reflexiones vertidas en su célebre obra de 1934, *Pensamiento y Lenguaje*.

En el núcleo de este programa se halla la idea de que, para comprender el “desarrollo cultural” de los *miembros* de nuestra especie, debemos de separarlo del estudio del nuestro desarrollo filogenético: mientras que en este segundo caso buscamos determinar los mecanismos cognitivos modelado por la naturaleza para *hacer frente al medio*, en el primero estamos interesados en comprender cómo nos es posible incorporar a nuestra estructura mental los desarrollos culturales, principalmente de naturaleza simbólica, que nos permiten *transformar nuestro medio*. Esta idea queda perfectamente abarcada por lo que se ha dado en llamar la “ley general del desarrollo cultural” (Vygotsky, 1978, citado en Wertsch, 1988, p. 60):

Cada función en el desarrollo cultural del niño aparece dos veces o sobre dos niveles distintos. Primero aparece en el nivel social y, entonces, tiene lugar en el nivel psicológico. Primero aparece entre personas como una categoría interpsicológica y posteriormente dentro del niño como una categoría intrapsicológica. [...] Las relaciones sociales o relaciones entre personas están genéticamente en la base de todas las funciones superiores y sus relaciones.

Debe quedar claro que Vygotsky entiende que las “funciones superiores” son un conjunto de procesos psicológicos exclusivos de nuestra especie (en pensamiento matemático, la atención voluntaria, la memoria lógica, la formación de conceptos, el desarrollo de la volición, por mencionar sólo algunos) que comparten la particularidad de estar *mediados* por *sistemas*

simbólicos culturalmente desarrollados (comprensibles sólo a partir del conocimiento de la evolución histórica de sociedades particulares) que son *interiorizados* a partir de su uso en *contextos dialógicos*, hasta terminar formado parte del *proceso privado de pensamiento*.

Por otra parte, Vygotsky no cae en el error ingenuo de negar que la nuestra cognición tenga una sólida base biológica; por el contrario, pone gran esmero en intentar mostrar que hasta los procesos que consideramos más típicamente humanos, como lo son el razonamiento y el lenguaje, están en diferente grado presentes a lo largo del escalar filogenético. El poder de la cultura no está en su *posibilidad* de replicarse en mentes particulares, sino en su *capacidad* de poner en contacto competencias que, hasta entonces, habían seguido caminos independientes a lo largo de nuestra evolución como especie. Sin embargo, el papel de la cultura en el desarrollo cognitivo está más allá de continuar el camino trazado por la biología; refiriéndose a la relación entre pensamiento y habla, Vygotsky (1934/1986) afirma:

Si comparamos el desarrollo temprano del habla y el intelecto –que, como hemos visto, se desarrollan a través de líneas separadas tanto en animales como en niños muy pequeños– con el desarrollo del habla interna y el pensamiento verbal, deberíamos concluir que la etapa posterior no es simplemente una continuación de la primera. *La propia naturaleza del desarrollo cambia*, desde lo biológico hacia lo socio-cultural. El pensamiento verbal no es una forma de conducta natural e innata, sino que está determinada por procesos histórico-culturales y tiene propiedades y leyes específicas que no pueden ser encontradas en las formas naturales del pensamiento y el habla (p. 94).

Así pues, en su explicación del desarrollo cultural infantil, Vygotsky (1934/1986) otorga una importancia preponderante al proceso de interiorización del lenguaje y a su consecuente impacto sobre la estructura global de la cognición. Él ve al lenguaje como la *herramienta intelectual* por excelencia, que en un primer momento entra en juego en las interacciones comunicativas entre el niño y sus cuidadores por medio del “habla externa” para posteriormente convertirse en “habla interna”, guiando así el propio proceso privado de pensamiento. El sólo a partir de la cultura que puede ocurrir el proceso de adquisición del lenguaje, pero es sólo a través de la apropiación del lenguaje que puede tener lugar una participación plena en la cultura. De este modo, el “pensamiento verbal” (que, a la postre, sería el único género de pensamiento genuinamente interesante desde el punto de vista psicológico) es entendido como un tipo de “charla interna” que posibilita la adquisición y uso de otras “herramientas intelectuales” de naturaleza eminentemente social, que van desde la memoria lógica y la formación de conceptos hasta la propia conciencia que cada uno de nosotros tiene de sí mismo.

Resulta complicado dimensionar la influencia que este tipo de argumentos tiene sobre la teorización psicológica general. La aceptación de su premisa básica (a decir, que pensamos a partir de nuestra lengua materna y que, por lo tanto, el pensar presupone el dominio del lenguaje) resulta tan intuitiva que durante mucho tiempo se pensó que ni siquiera tenía que ser argumentada. Ideas similares pueden ser encontradas en el trabajo de otros influyentes teóricos como Jean Piaget (1928/2002), quien sostiene que:

La necesidad social de compartir los pensamientos de los otros y de comunicar los propios con éxito está en la base de nuestra necesidad de verificación. [...] El razonamiento lógico es un argumento que tenemos con nosotros mismos y que reproduce internamente las características de un argumento real. [...] La conversación discursiva y el razonamiento son el producto de la interacción entre individuos (p. 204).

En suma, nos hallamos frente a una perspectiva que: (1) enfatiza la dimensión social del lenguaje, (2) sugiere que la adquisición del lenguaje *transforma*, por no decir que posibilita, el proceso de pensamiento y (3) sostiene que el razonamiento *replica* una interacción dialógica (*i.e.*, social), donde el propio sujeto es a la vez quien propone y evalúa sus razones.

He destacado el carácter intuitivo de este tipo de argumentos; no obstante, una de sus principales limitaciones recae sobre su imposibilidad de dar una respuesta convincente a una observación que resulta igualmente atractiva desde el punto de vista del sentido común: si el pensamiento presupone el dominio del lenguaje, entonces los organismos que carecen de lenguaje habrían de carecer igualmente de pensamiento. Sin embargo, hay algo en nuestra experiencia cotidiana que nos hace pensar que una parte importante de la conducta emitida por los animales no humanos y por los niños preverbales es demasiado sofisticada para ser explicada *exclusivamente* en términos de instintos, ensayo y error o mecanismos de aprendizaje asociativo (Bermúdez, 2003; Penn y Povinelli, 2007). Por el contrario, la comprensión de ciertos comportamientos presentes en los organismos no-

verbales parece requerir la apelación a procesos de razonamiento *análogos* a los atribuibles a un humano lingüísticamente competente.

Sin embargo, esto no parece tener ningún sentido si hacemos caso a nuestra experiencia fenoménica que nos indica que nuestro pensamiento se estructura *a partir* de nuestro lenguaje. Nos hallamos ante la confrontación abierta de dos de nuestras intuiciones más arraigadas, ¿hay manera de librar esta contradicción?

2.1.2. EL ARGUMENTO SOFISTICADO (AS) SOBRE LA NATURALEZA LINGÜÍSTICA DE LA COGNICIÓN.

Las ideas que englobo bajo la etiqueta de “Argumento Sofisticado” (AS) típicamente se identifican como los precursores directos del desarrollo de la ciencia cognitiva. A diferencia del AI, que sostiene que el pensamiento hereda su estructura de una lengua natural, en el AS se entiende que la naturaleza lingüística *de la cognición en general* deviene de su estatus como un fenómeno esencialmente *computacional* y que, por tanto, es previa a la adquisición de cualquier sistema simbólico culturalmente desarrollado.

Aquí los paralelos entre el estudio de la cognición y las discusiones en filosofía del lenguaje y la ciencia se vuelven más diversos pero también más claros. Tres aspectos son los que me gustaría destacar: (1) el supuesto de que las teorías científicas son construcciones lingüísticas tales que, una teoría, T1, puede ser *reducida* (traducidos sin pérdida de contenido) al lenguaje de otra teoría, T2, de carácter más general, donde esta reducción

de T1 a T2 puede resultar útil para disolver los problemas conceptuales inherentes a T1; (2) el impacto que el progreso de la filosofía del lenguaje y la lógica tuvo sobre la evolución de disciplinas como las matemáticas y la informática, y (3) la subsecuente extensión de estos desarrollos a la ciencia cognitiva. Estas tres consideraciones son de vital importancia para comprender el desarrollo histórico del AS (véase Gardner, 1985, para una exposición detallada del desarrollo de la ciencia cognitiva).

En su trabajo sobre filosofía de las matemáticas, Bertrand Russell (influido profundamente por el pensamiento de Frege) se da a la tarea de resolver algunos de los problemas inherentes a la teoría de los números, tratando a éstos como *construcciones lógicas a partir de conjuntos* (citado en Putnam, 1963/1983). Esta reducción de la teoría de los números a la teoría de conjuntos a partir de la lógica constituye la primera gran “victoria” (si no es que acaso la única) de la filosofía analítica, y es un éxito que caló hondo en el desarrollo de la filosofía y la ciencia en su totalidad. Entender su impacto sobre el estudio de la cognición y el comportamiento es de suma importancia para comprender la evolución de la ciencia psicológica a lo largo de todo el siglo XX.

Recordemos que en el capítulo anterior argumenté que el conductismo lógico, emanado directamente de las consideraciones de los miembros del Círculo de Viena sobre la naturaleza de los fenómenos psicológicos (Carnap, 1932/2002; Hempel, 1931/1980), sirvió de base conceptual para la tradición conductista en psicología defendida por autores como John B. Watson (1930/1970) y Burrhus Frederic Skinner (1938/1975). En el núcleo

de las consideraciones conductistas (tanto en psicología como en filosofía) se halla la idea de que el lenguaje mentalista (plagado de conceptos más allá de lo empíricamente verificable) constituye un problema para nuestra comprensión del comportamiento, por lo que habría de ser reducido al lenguaje extensional propio de la descripción de la conducta manifiesta. Hilary Putnam (1963/1983) es muy claro al identificar la relación entre las propuestas de Russell y las pretensiones del empirismo lógico, donde el conductismo es sólo una expresión más. Al respecto escribe:

Lo novedoso de esto fue la idea de deshacerse de ciertas entidades filosóficamente indeseables o embarazosas (los números) sin dejar de hacer justicia a la adecuada configuración del discurso (la teoría de los números) tratando a las aludidas entidades como construcciones lógicas. Russell se apresuró a erigir este “éxito” como modelo para todos los filósofos futuros. Y algunos de estos filósofos futuros —los positivistas de Viena en su fase “fiscalista” (alrededor de 1930) —tomaron tan en serio el consejo de Russell que idearon la doctrina que llamamos *conductismo lógico*, según la cual, del mismo modo que los números son (según se pretende) construcciones lógicas a partir de conjuntos, los *sucesos mentales* son construcciones lógicas a partir de *sucesos* de conducta reales y posibles (p. 6).

Este acoso reduccionista sobre el lenguaje mentalista va más allá de cualquier forma de conductismo. Autores como J.J.C Smart (1959), David K. Lewis (1966/1984) o David M. Armstrong (1970/1980) —reconociendo las limitaciones intrínsecas a los posicionamientos conductistas— definden la idea de que al emplear términos mentales en realidad nos estamos

refiriendo a configuraciones fisicoquímicas del cerebro, por lo que el lenguaje psicológico podría ser sustituido por el lenguaje neurofisiológico. La idea de que los estados mentales son *idénticos* a estados cerebrales está aún muy presente en el espíritu de la neurociencia contemporánea.

De este modo, si la psicología aspiraba a erigirse como una ciencia independiente, al margen de cualquier intento de reduccionismo, había de encontrar una forma hacer compatible el lenguaje mentalista (las explicaciones psicológicas en términos de deseo-creencia) con los principios mecanicistas del naturalismo; como Jerry Fodor (1975) comenta, se puede ser *fisicalista* sin ser *reduccionista*. Paradójicamente, la respuesta que garantizará la autonomía de la explicación psicológica comparte sus raíces teóricas con la tradición reduccionista en filosofía.

Es necesario tener presente que, como lo comenta Zenon W. Pylyshyn (1989/2007), el trabajo de Frege y Russell es fundamental para comprender el desarrollo de la lógica matemática y, en consecuencia, el florecimiento de disciplinas como la informática y la computación. El precisamente Alan Turing, uno de los teóricos más destacados en estos campos del saber, quien sentará las bases sobre las que se gestará y progresará la llamada “revolución cognitiva”. Turing (1936), en un trabajo donde aborda el problema de la decibilidad de las matemáticas, plantea que “es posible construir una máquina simple para computar cualquier secuencia computable” (p. 241). La idea es que es plausible describir abstractamente una “máquina universal” (*universal computing machine*), MU, que sea capaz de *simular* cualquier otro ordenador *formalmente especificable*.

En su ya clásico artículo de 1950, “Computing Machinery and Intelligence”, Turing va un paso más allá al proponer un criterio objetivo para atribuir la capacidad de pensamiento, lo que él llama el “juego de la imitación” y que posteriormente sería conocido como “la prueba de Turing”. El juego consiste en que un interrogador (A) que formula preguntas a una persona (B) y a un ordenador (C) que se hallan aislados en otra habitación; el formato de la presentación de las respuestas de B y C se homogeneiza (por ejemplo, las respuestas se presentan de forma escrita) de modo tal que A no pueda distinguir entre ambos sólo a partir de características superficiales de su respuesta (como el tono de voz, por ejemplo). La idea de Turing es que si después del interrogatorio A no puede determinar cuál de los interrogados era la persona y cuál el ordenador, ¿no hay razón alguna para negar la adscripción de pensamiento a la computadora!

Con su trabajo, Turing no sólo inaugura la investigación en el campo de la Inteligencia Artificial (IA), sino que también revoluciona la forma de comprender y abordar la *inteligencia biológica*. Hay que distinguir dos elementos principales en movimiento de Turing: (1) él equipara los procesos computacionales con los procesos mentales (*i.e.* hace que el “comportamiento” de los ordenadores sea susceptible de explicación psicológica) y (2) conceptualiza la mente (de forma particular, la mente humana) como uno más entre los múltiples ordenadores sujeto a ser simulados por la máquina universal: nace, pues, la metáfora del ordenador y con ella la *teoría computacional de la mente*.

Es importante trascender las interpretaciones triviales de la metáfora del ordenador para comprender plenamente la magnitud del cambio que representó para la teorización en torno a los fenómenos mentales. En la actualidad (en que la capacidad y sofisticación de las computadoras personales se duplica año con año) resulta muy sencillo establecer analogías superficiales entre éstas y la mente/cerebro en términos de sus *componentes*: ambas parecen tener una Unidad Central de Procesamiento (CPU), sistemas de input y de output, unidades dedicadas de memoria, etc. Sin embargo, en el corazón de la metáfora hay una idea con implicaciones conceptuales mucho más profundas e interesantes. Esto es, tanto las computadoras, como la mente, puede ser descritos como *sistemas simbólicos*: la metáfora del ordenador es, ante todo, la metáfora de la mente como un lenguaje formal. Preguntarse por los “componentes” de la mente (por ejemplo, sobre la capacidad de la memoria) sólo tiene sentido en la medida en que se acepta que ésta es, ante todo, una entidad lingüística, constituida por un conjunto *símbolos* (un vocabulario) y otro conjunto de *reglas de formación, transformación e inferencia*. Esta idea queda más claramente expresada por Zenon W. Pylyshyn (1989/2007), quien afirma que:

En computación científica [...] las transformaciones sobre expresiones más comunes eran aquellas que reflejan funciones matemáticas sobre números. Pero si las expresiones simbólicas codificaran proposiciones o creencias o conocimiento, tal y como podría ocurrir si fueran expresiones en alguna lógica simbólica, entonces el ordenador podría transformarlas en formas correspondientes a pruebas o inferencias, o quizás una secuencia de «pensamientos» que tienen lugar durante el

razonamiento ordinario. Lo importante es que, de acuerdo a la visión clásica, ciertos tipos de sistema, incluidos tanto mentes como ordenadores, operan sobre *representaciones* que toman la forma de códigos simbólicos (p 30).

Esto bloquea cualquier tipo de reduccionismo del lenguaje psicológico al hacer ver que: (1) entre la estimulación ambiental y la emisión de conducta deben de ocurrir una serie de cálculos que permitan transitar entre un estado y otro (las ciencias cognitivas “abren la caja negra” ignorada por todas las formas de conductismo), y (2) al mostrar que, en tanto fenómenos computacionales, los fenómenos mentales pueden ser instanciados por diferentes soportes físicos (por ejemplo, conjuntos de neuronas o placas de silicio), se cierra la puerta a las forma más radicales de la teoría de la identidad (Putnam, 1967/1981; Fodor, 1975).

Tal cual está planteado hasta este punto, el *funcionalismo psicológico* defendido por autores como Putnam, Fodor y Pylyshyn, ya es en sí mismo una teoría lingüística de la *arquitectura mental*. No obstante, Fodor (1975) y Fodor y Pylyshyn (1988) van más allá al argumentar que, aunque la computación ya es en sí misma una noción lingüística, ésta presupone a su vez la existencia de un “medio en el cual se computa”, un sistema representacional que equiparan plenamente con un lenguaje: *el lenguaje del pensamiento*.

Fodor y Pylyshyn (1988) —en un artículo donde dan respuestas a las críticas expuestas por los teóricos conexionistas al modelo representacional clásico— argumentan que la naturaleza composicional, sistemática y

coherente de los procesos cognitivos sólo puede ser explicada a partir del supuesto de que las representaciones mentales tienen estructura interna (*i.e.*, de que existe un lenguaje interno, previo al aprendizaje de cualquier sistema simbólico creado por la cultura).

Recordemos que en el AI el lenguaje es visto como una herramienta eminentemente social que al ser *interiorizada* transforma la naturaleza del pensamiento. En el caso del AS, particularmente en la hipótesis del lenguaje del pensamiento, la adquisición de un lenguaje *público* como sistema de comunicación es vista más como la consecuencia que como la causa del carácter estructurado y sistemático de la cognición. Profundamente influenciado por el trabajo de Noam Chomsky (1957/2002; 1965) sobre gramática generativa, Fodor (1975) sostiene que:

La refutación más obvia (y, yo pensaría, suficiente) del supuesto de que los lenguajes naturales son el medio a través del cual pensamos es que hay organismos no verbales que piensan. [...] Los modelos computacionales presuponen sistemas representacionales. Pero los sistemas representacionales de los organismos preverbales e infrahumanos seguramente no pueden ser los lenguajes naturales (p. 56).

Este movimiento no es para nada menor. Al postular la existencia de un código interno como la esencia de la cognición en tanto fenómeno natural, Fodor (1975) no sólo diluye el embarazoso problema, presente en el AI, de la continuidad entre el pensamiento animal y el pensamiento humano, sino que también sienta las bases de un esquema general de explicación para los fenómenos mentales.

De este modo, todo suceso psicológico interesante desde el punto de vista de las ciencias cognitivas podría ser explicado a partir de lo que Bermúdez (2003) llama el “modelo teórico de decisión del razonamiento práctico” (*decision-theoretic model of practical reasoning*) (véase Fodor, 1975, pp. 28-29, para una descripción más detallada de la propuesta), entendido como un proceso de formación y confirmación de hipótesis. Como el mismo Fodor (1975) comenta:

Enfaticé, con propósitos expositivos, la significancia de la representación que el organismo tiene de su propia conducta en la explicación sus acciones deliberadas. Pero, una vez hecho, el argumento resulta ser mucho más oblicuo. Por ejemplo, está implícito en el modelo que el organismo dispone de medios para representar no sólo sus opciones conductuales, sino también: las probables consecuencias de actuar sobre esas opciones, un ordenamiento de preferencias definidas sobre esas consecuencias y, por supuesto, la situación original en que él mismo se encuentra. Usar este tipo de modelo es, por lo tanto, presuponer que el agente tiene acceso a un sistema representacional de una riqueza bastante considerable. De acuerdo con el modelo, la toma de decisiones es un proceso computacional; el acto que el agente lleva a cabo es la consecuencia de computaciones definidas sobre representaciones de posibles acciones. Sin representación no hay computación. Sin computación no hay modelo. [...] También pude haber dicho que el modelo presupone un lenguaje (p. 31).

De acuerdo con Fodor, es sólo a partir de estos supuestos que se puede dar cuenta de fenómenos psicológicos tan prominentes como la percepción, la toma de decisiones y el aprendizaje de conceptos (*i.e.*, de la propia adquisi-

ción del lenguaje). De este modo, el AS a favor de la naturaleza lingüística de la cognición desdibuja muchas de las dicotomías propias de la investigación psicológica: la distinción entre teorías normativas y teorías descriptivas, entre cognición humana y cognición animal, entre lenguajes formales y lenguajes naturales, entre racionalidad teórica y racionalidad práctica, por mencionar sólo algunas. A partir de la llamada revolución cognitiva, resulta muy complicado distinguir entre las teorías filosóficas y científicas sobre el lenguaje de aquellas que pretenden caracterizar la mente.

2.2. LAS DOS METÁFORAS LINGÜÍSTICAS DE LA MENTE: EL LIBRO Y LA COMPUTADORA.

Para favorecer la claridad en torno a las diferencias existentes entre el AI y el AS, resulta útil contextualizar la discusión en un marco más general de las disputas entre teorías estructuralistas y teorías funcionalistas que ha tenido lugar dentro de las ciencias sociales. Con esto pretendo dejar aún más de manifiesto el espíritu lingüístico detrás de una y otra línea de argumentación.

Con lo dicho hasta ahora, debería quedar más que claro que el AS es típico de las teorías funcionalistas, las cuales parten de una visión analítica de los fenómenos orientada a la descripción de los mecanismos que subyacen a su ocurrencia. Esta es la posición defendida por Putnam, Fodor y Pylyshyn en lo que tiene que ver con la naturaleza de los fenómenos mentales. No obstante, resulta importante regresar un poco sobre nuestros

pasos para considerar con más detalle la relación entre el AI y las teorías estructuralistas, así como la profundidad de sus implicaciones teóricas.

Es justo decir que, empero sus importantes aportaciones a la investigación experimental y empírica de la cognición y el comportamiento, el programa vygotskyano debe ser entendido en el contexto más amplio de la teoría educativa constructivista y de las perspectivas narrativas del estudio de la mente. A pesar de su heterogeneidad, todas estas propuestas pueden ser englobadas en una misma tradición intelectual, el estructuralismo, cuyos fundamentos se remontan a los trabajos de autores como Ferdinand de Saussure (1916/2008) en lingüística y Claude Lévi-Strauss (1973/1979) en antropología, así como al Wittgenstein (1953/1997) de las *Investigaciones Filosóficas*.

Estas consideraciones son importantes dado que se ha supuesto que esta “aproximación narrativa” constituye una alternativa a la visión sintáctica y formalista de la mente propia del AS. No obstante, y aun aceptando un genuino antagonismo entre estas dos posiciones, me gustaría dejar claro que las críticas aquí expuestas aplican por igual a ambas posturas.

Lo que diferencia a una aproximación de la otra no es el grado en que cada una toma al lenguaje como un modelo para caracterizar los fenómenos psicológicos, sino *la caracterización del lenguaje* que una y otra considera más adecuada para estudiar la mente: mientras que los partidarios de la visión narrativa enarbolan la visión saussuriana del lenguaje (de Saussure, 1916/2008), los teóricos funcionalistas hacen lo propio con la caracterización chomskiana (Chomsky, 1957/2002; 1965).

Esta idea puede quedar aún más clara si consideramos las metáforas de las que echan mano los teóricos de cada tradición en sus esfuerzos por esclarecer la naturaleza de los fenómenos mentales. Como comenté más arriba, la metáfora por excelencia de las argumentaciones sofisticadas a favor del modelo del lenguaje es el ordenador digital: se enfatiza el carácter simbólico de los procesos cognitivos, sosteniendo que es posible definirlos en términos sintácticos (*i.e.*, computacionales) como la aplicación de un conjunto de reglas sobre representaciones discretas.

Por su parte, los defensores de la aproximación narrativa hacen uso de una de una analogía radicalmente diferente que, no por ello, resulta menos lingüística: el libro. De este modo, Paul Ricoeur (1970/1986; 1971/1986; 1977/1986) sostiene que es conveniente tomar *al texto* como un modelo para caracterizar y comprender la “acción significativa”, la cual sería el objeto natural de estudio de las ciencias humanas. Esta idea, tan extendida en disciplinas como la antropología y la sociología, tiene una de sus expresiones más claras e influyentes dentro de la psicología en el trabajo de Jerome Bruner (1991), quien afirma que “[la psicología popular] es de naturaleza narrativa en lugar de lógica o categórica” (p.55). Es decir, Bruner sostiene que nuestro comportamiento cotidiano no responde tanto a la aplicación de un conjunto preestablecido de reglas formales sobre representaciones discretas de los eventos, sino que se organiza en torno a una trama novelesca de la cual somos tanto partícipes como narradores.

Lo interesante de la perspectiva estructuralista es que la naturaleza lingüística de la mente (su propiedad de estar organizada como una narrativa)

es a la vez una causa y una consecuencia de la naturaleza lingüística de las *interacciones entre mentes*. En este sentido, de Saussure (1916/2008) ya prefiguraba: “una ciencia que estudie la vida de los signos en el seno de la vida social; formaría parte de la psicología social, y, por consiguiente, de la psicología general; la denominaremos *semiología* (del griego *sémeion*, “signo”).” (pp. 42-43). La manifestación más clara de esta semiología es el programa estructuralista en antropología inaugurado por Lévi-Strauss (1973/1979), en cuyo centro se halla la idea de que los intercambios sociales en todos los niveles están regidos por una serie de *códigos tácitos* y que la labor del antropólogo está precisamente en hacer explícitas las *relaciones* existentes entre los diversos elementos de la cultura para de esta forma poder reconstruir su *significado*.

En este capítulo presenté las que, considero, son las líneas de argumentación más importantes que subyacen a concepción lingüística de la mente —el Argumento Intuitivo (AI) y el Argumento Sofisticado (AS) a favor del “modelo del lenguaje”—, así como su relación con las tradiciones intelectuales que típicamente las sustentan: el estructuralismo, con el AI, y el funcionalismo, con el AS.

Es importante hacer ver que la correspondencia entre tradiciones intelectuales y líneas de argumentación no es perfecta (como quedó claro al revisar la propuesta de Frege y como se verá al revisar la postulación de “teorías implícitas” por parte de los psicólogos del desarrollo), sin embargo, sí existen una tendencia muy marcada que permite vincular las unas con

las otras, y no sólo con fines expositivos, sino que, sobre todo, para entender las disputas teóricas existentes entre funcionalismo y estructuralismo.

Así pues, resulta evidente la vena lingüística detrás de estas dos influentes tradiciones intelectuales. Con todo esto, habría de quedar claro que cuando me refiero al “modelo del lenguaje” no me estoy restringiendo al hecho, casi trivial, de que nuestra lengua materna es el vehículo a través del cual fluye nuestro pensamiento. La concepción simbólica de la consciencia que tenemos de nosotros mismos está a la mitad entre dos líneas de argumentación con implicaciones teóricas muchísimo más profundas: el AI que, en última instancia, sostiene que las interacciones sociales *son estructuras lingüísticas*, y el AS que hace de todo fenómeno mental un fenómeno computacional. En los dos capítulos siguientes abordo con más profundidad las implicaciones del AS para la investigación en ciencia cognitiva.

CAPÍTULO TRES

EL AS EN LA CIENCIA COGNITIVA CONTEMPORÁNEA: DOMINIO ESPECIFICIDAD Y MODELOS MATEMÁTICOS

I take it that computational processes are both symbolic and formal. They are symbolic because they are defined over representations, and they are formal because they apply to representations, in virtue of (roughly) the syntax of the representations.

Jerry Fodor

La revisión del AI y el AS, así como de las tradiciones intelectuales que los sustentan, permitirá comprender de mucha mejor forma el estado actual de la investigación y teorización en ciencias cognitivas, objetivo que persigo en los dos siguientes capítulos. A grandes rasgos, a partir de los años '80 del siglo pasado, el estudio empírico de los fenómenos mentales ha tendido a agruparse en torno a dos programas de investigación principales: (1) las teorías modularistas y de dominio especificidad, por un lado (capítulo 3), y (2) el análisis experimental de la conducta, particularmente a través del uso de modelos matemáticos para describir los procesos cognitivos, por el otro (capítulo 4).

Tradicionalmente, se ha venido defendiendo la idea de que hay razones de fondo para suponer que ambas posturas son excluyentes, lo que ha promovido el surgimiento de ásperas disputas teóricas entre los defensores de

una y otra posición. El propósito de los siguientes dos capítulos es mostrar que este supuesto está por demás infundado, dato que, más que alternativas a la conceptualización de los procesos cognitivos, ambas posturas representan dos caras de una misma moneda: la concepción sintáctica de los fenómenos mentales. La diferencia está en el énfasis formalista de cada posición, lo cual me lleva a afirmar ambas pueden coexistir perfectamente.

3.1. MÓDULOS MENTALES Y TEORÍAS IMPLÍCITAS: LA ESPECIFICIDAD DE DOMINIO Y SUS IMPLICACIONES PARA LA COMPRENSIÓN DE LA ARQUITECTURA MENTAL

Como muy bien lo comentan Ángeles Eraña (2006) y Josep Maria Domingo (2003), no siempre queda claro a qué se están refiriendo los científicos cognitivos cuando hablan de sistemas de conocimiento específicos de dominio. Es necesario, de entrada, diferenciar entre *módulos mentales* y *teorías intuitivas*, así como establecer la relación entre unos y otras.

Desmarañar estas dos complicadas nociones demandan revisar con más detalle lo que dos de los autores más influyentes de la tradición computacional clásica tienen que decir al respecto: Noam Chomsky (1957/2002; 1965), en sus trabajos pioneros sobre *gramática generativa*, y, una vez más, Jerry Fodor (1983), con su propuesta de la estructura modular de los sistemas mentales periféricos.

3.1.1. LA ORGANIZACIÓN MODULAR DE LA FACULTAD DE LENGUAJE.

Hablar de cognición sin hablar de Noam Chomsky resultaría una omisión casi imperdonable. Con su trabajo sobre la competencia lingüística de nuestra especie, Chomsky (1957/2002; 1965) marca el camino a seguir por resto de los interesados en la caracterización y estudio de los fenómenos mentales. No es gratuito que el padre de la lingüística moderna sea a la vez uno de los personajes más importantes de la llamada “revolución cognitiva” de mediados del siglo XX. Los comentarios de Chomsky (1959) a *Verbal Behavior* de Skinner (1957) representan un hito en la transición de la hegemonía del conductismo hacia las teorías computacionales del funcionamiento mental.

Chomsky (1957/2002; 1959; 1965) importa al campo del estudio de las lenguas naturales la caracterización desarrollada por autores como Frege (véase Moravcsik, 2008, para un análisis de las similitudes entre Frege y Chomsky) y Carnap (1937) en el contexto de la filosofía formalista del lenguaje. La idea clave de Chomsky (1957/2002; 1959; 1965) es que sólo se puede dar cuenta de la adquisición de una lengua a partir de la postulación de un *conocimiento lingüístico previo a la experiencia*, una “gramática universal” innata común a todos los lenguajes humanos.

En *Aspects of the theory of syntax*, Chomsky (1965) propone la existencia de un “dispositivo para la adquisición del lenguaje” (*language-acquisition device*) como el sustento físico de la “gramática universal” y sienta las bases para extender su propuesta a toda la arquitectura cognitiva:

El dispositivo para la adquisición del lenguaje es sólo un componente del sistema total de estructuras intelectuales que pueden ser aplicadas a la solución de problemas y la formación de conceptos; en otras palabras, la *faculté de langage* es sólo una de las facultades de la mente. [...] El problema de mapear las capacidades cognitivas intrínsecas de un organismo, e identificar los sistemas de creencia y la organización conductual que éstos pueden alcanzar fácilmente, debería ser central en la psicología experimental (Chomsky, 1965, pp. 56-57).

El camino trazado por Chomsky resultó tan atractivo que muchos optaron por equiparlo con la totalidad de la ciencia cognitiva. De este modo, autores como Marc D. Hauser (2006) tienen a bien considerar que esquema chomskiano de explicación puede resultar útil para dar cuenta de otros aspectos clave de nuestra vida mental. En su libro *Moral Minds: The nature of right and wrong*, Hauser (2006) afirma que la dinámica de nuestros juicios sobre el bien y el mal puede ser capturada a través de la postulación de una “gramática moral”, un sistema de reglas que opera sobre la información emanada de las interacciones entre personas:

Para entender nuestra psicología moral, no exploraré todas las formas en que hacemos uso de ella en nuestras interacciones cotidianas con otras personas. Del mismo modo en que los lingüistas de la tradición chomskiana pasan por alto las cuestiones concernientes al uso del lenguaje para concentrarse en el conocimiento inconsciente que permite a cada uno de nosotros expresar y evaluar un número ilimitado de sentencias, yo adopto una aproximación igualmente restringida con respecto a la moralidad. El resultado es una explicación rica y detallada

de cómo es que una gramática moral inconsciente y universal subyace a nuestros juicios sobre el bien y el mal (p. viii).

El hecho de que Hauser (2006) no dude en postular una “gramática moral” es elocuente en muchos sentidos. Primero, la caracterización chomskiana del lenguaje nos muestra de forma clara que a cincuenta años de distancia de que fue originalmente formulada aún tiene mucho que dar de sí. Aún más importantes son las implicaciones teóricas inherentes a esta extrapolación del modelo del lenguaje; se acepta que: (1) los fenómenos mentales son, en esencia, de naturaleza sintáctica, (2) el conocimiento de las *reglas* a partir de las que se estructura nuestra percepción y comprensión del mundo (lingüístico, social, moral, físico, etc.) es innato, (3) muy vinculado con el punto anterior, los sistemas de creencias que podemos construir a partir de nuestra experiencia están fuertemente constreñidos por nuestro conocimiento innato (dado que éste determina nuestra propia experiencia), (4) las diferencias entre individuos y entre culturas son sólo superficiales y enmascaran patrones universales de razonamiento, (5) no existen reglas básicas de aprendizaje, sino una serie de “gramáticas” específicas de dominio que nos permiten navegar distintas parcelas de la realidad, y (6) nuestra comprensión del mundo no es unitaria, sino que se estructura a partir de una serie de cuerpos nucleares de conocimiento definidos en términos de la información ambiental que los constituye y las reglas que sobre ella se aplican (aunque, véase Fodor, 1983 y la sección siguiente).

3.1.2. LA ARQUITECTURA MODULAR DE LOS SISTEMAS MENTALES PERIFÉRICOS

En su influyente trabajo, *The Modularity of Mind*, Fodor (1983) establece una distinción entre sistemas cognitivos periféricos, que son *módulos* computacionales “específicos de dominio, innatamente especificados, compactos, autónomos y no ensamblados” (p. 37), y sistemas cognitivos centrales, que, básicamente, tendrían las propiedades contrarias. Fodor reconoce que los mayores avances en IA son los que han tenido lugar en el modelamiento de procesos cognitivos relativamente sencillos, como la percepción y el lenguaje, mientras que la “psicología del pensamiento” ha permanecido, por su complejidad, como un problema prácticamente intratable. Nos hallamos aquí ante la necesidad de diferenciar entre “computaciones relativamente locales y computaciones relativamente globales” (p. 126).

En tanto mecanismos dedicados a la resolución de problemas computacionales específicos, los sistemas periféricos están sujetos al tipo de descripciones funcionales características de la teoría representacional clásica. A grandes rasgos, todos estos módulos compartirían la propiedad de estar *encapsulados*; es decir que tanto la información que a ellos ingresa, como los cálculos que sobre ella se aplican y las representaciones que de ellos resultan, serían específicos de cada módulo y prácticamente no interactuarían con los procesos que tienen lugar en los módulos adyacentes.

No obstante, con los sistemas centrales ocurre algo muy distinto y un tanto perturbador. Fodor (1983) considera que una arquitectura mental puramente modular sería simplemente incompatible con la flexibilidad cog-

nitiva exhibida por la mente humana —manifiesta en procesos como el control ejecutivo, la fijación de creencias, la elección de hipótesis, la inteligencia general, la creatividad abductiva o el razonamiento analógico. Sin embargo, dado que la teoría computacional clásica es la única aproximación fructífera al estudio de los fenómenos psicológicos (al menos para aquellos que la defienden, como es claro en el caso de Fodor), postular la existencia de sistemas cognitivos no modulares equivale a aceptar que éstos ¡serían inexplicables desde el punto de vista psicológico!

En una clara analogía con la filosofía de la ciencia, Fodor (1983) proclama que los sistemas cognitivos centrales, al igual que la confirmación de las teorías científicas, son “isotrópicos” (es decir, que los hechos relevantes para la confirmación de una hipótesis pueden ser extraídos de cualquier parte en el campo de las verdades aceptadas como tales por un agente, p. 105) y “quineanos” (lo que implica que “el grado de confirmación asignado a una hipótesis determinada es sensible a las propiedades de sistema global de creencias”, p. 107). Así pues, queda claro que los sistemas centrales no pueden ser específicos de un dominio, sino que son *facultades horizontales* que tienen acceso libre a la totalidad de los datos disponibles. Es precisamente este holismo el que sitúa a los sistemas centrales más allá de abordaje empírico posibilitado por la teoría simbólico-representacional clásica. Aunque ciertamente es una estrategia arriesgada, Fodor (1983) prefiere limitar en buena medida las perspectivas de progreso de las ciencias cognitivas a ser simplemente omiso de las propiedades que resultan más interesantes en el proceso de pensamiento:

Sugiero que la razón por la que no hay una psicología seria de los procesos cognitivos centrales es la misma razón por la que no hay una filosofía seria de la confirmación científica. Ambos casos ejemplifican la significancia de factores globales en la fijación de creencias y nadie ha siquiera comenzado a comprender cómo tales factores ejercen sus efectos. En este sentido, la ciencia cognitiva ni siquiera ha *comenzado*; literalmente, no hemos avanzado más lejos de los que estábamos en los oscuros días del conductismo (aunque estamos, no cabe duda, más desilusionados en algún sentido benéfico) (p. 129).

Con su noción de módulo, Fodor (1983) ofrece una caracterización sumamente útil a los interesados en el estudio de los procesos mentales y sus sustratos neurofisiológicos. Y, aún más importante, plantea un reto que desde entonces se convirtió en la motivación primordial de la ciencia cognitiva: encontrar la mejor forma de afrontar el estudio de los procesos centrales de pensamiento.

3.1.3. DESARROLLOS ACTUALES EN MODULARIDAD MENTAL Y DOMINIO ESPECIFICIDAD

Hasta aquí, resulta útil distinguir entre dos nociones básicas de “módulo” mental. Fodor (1983) establece una clara distinción entre: (1) los *sistemas de entrada (input systems)* —o lo que he llamado “sistemas periféricos”— y (2) los “módulos chomskianos”, entendidos como *sistemas de información* específica de dominio (véase también Eraña, 2006). Considero que

con lo dicho hasta el momento es suficiente para tener una idea clara de qué es un sistema de entrada, sin embargo, resulta útil abundar en lo que tiene que ver con la caracterización de los “módulos chomskianos”. Al respecto Fodor (1983) escribe:

[Lo] que Chomsky piensa que es innato es principalmente un cierto *cuerpo de información*: el niño, por así decirlo, “nace conociendo” ciertos hechos con respecto a las restricciones sobre los posibles lenguajes humanos. Es a partir de la integración de este conocimiento humano con el cuerpo de los “datos lingüísticos primarios” [...] que se explica la eventual asimilación de las capacidades lingüísticas maduras. [...] La idea de que lo que es innato tiene un contenido proposicional es así parte y parcela de cierta visión de la ontogenia de las capacidades mentales —es decir, que en el desarrollo cognitivo, lo que está endógenamente dado está computacionalmente desplegado. (pp. 4-5).

El propio Chomsky (2009) ha enmarcado su teoría lingüística dentro del neocartesianismo, cuya esencia es, en palabras de Fodor (1983): “asumir que la estructura mental debería ser explicada principalmente en referencia al contenido proposicional de los estados mentales” (p. 6).

Es importante tener presente que el uso de concepto “módulo” es mucho más extenso dentro de las ciencias cognitivas —que van desde la filosofía hasta las neurociencias— (Domingo, 2003), aunque es claro que la distinción entre “sistemas de entrada” y “módulos chomskianos” puede servir como un buen punto de partida para reconstruir la discusión general en torno a esta importante noción. Me gustaría dejar de manifiesto que han

pocas áreas de investigación donde reine tanto la confusión como es en el caso de la discusión en torno a la modularidad. Con esto en mente, reconozco que la reconstrucción teórica que aquí realizó está muy lejos de ser conclusiva.

3.1.3.1. Modularidad de los sistemas cognitivos centrales: la tesis de la modularidad masiva

A grandes rasgos, existe un acuerdo relativamente amplio en que los sistemas perceptuales son modulares en el sentido originalmente propuesto por Fodor (1983). Por otro lado, si la intratabilidad de los procesos centrales del pensamiento radica en su no-modularidad, entonces una solución obvia a este problema es negar este hecho y defender la existencia de “módulos centrales”, sistemas cognitivos específicos de dominio cuyo *input* y *output* sean *información conceptual*.

Es precisamente esta la estrategia asumida por autores como Dan Sperber (1994), quien defiende que la arquitectura mental es *masivamente modular*. Sperber (1994, p. 43) propone un caso imaginario para justificar su postura: piénsese en los organismos de una determinada especie, los “protorgs”, que corren el riesgo de ser aplastados por los elefantes. El peligro está señalado por la *coocurrencia* del sonido de los elefantes al aproximarse (R) y las vibraciones provocadas en el suelo por su andar (V). Para hacer frente a este *problema adaptativo*, la selección natural dotó a los protorgs de dos módulos, uno para detectar R y otro para detectar V. La

activación de módulo R o el módulo V desencadena automáticamente una conducta de huida, lo cual aumenta la probabilidad de supervivencia de los protorgs. No obstante, habría de ser claro que podría ocurrir R sin V (o viceversa), quedando abierta la posibilidad de que los protorgs escapen del lugar sin que sea realmente preciso (la situación sólo es realmente peligrosa cuando se activan *al mismo tiempo* el módulo R y el módulo V), lo cual implicaría un gasto innecesario de energía. Con el tiempo, aparece una nueva especie, los orgs, que, además de estar equipada con los sistemas perceptuales de los protorgs, posee un *sistema de inferencia conceptual* (C) que recibe la información sobre activación de los módulos R y V y desencadena la conducta de huida *sólo cuando ambos son estimulados al mismo tiempo*. La lógica de la teoría evolutiva indica que, dejando de lado otros problemas adaptativos, con el tiempo protorgs se extinguiría y los orgs ocuparían su lugar.

Aún más importante, Sperber (1994) sostiene que:

El *input* de los primeros módulos conceptuales que aparecieron en la evolución de la cognición debió de haber surgido de los módulos perceptuales. Sin embargo, una vez que algunos módulos conceptuales se consolidaron en su lugar, sus *outputs* podrían haber servido como *inputs* de otros módulos conceptuales (p. 46).

Se entiende que, en la medida en que aumente la complejidad de las interconexiones entre diferentes sistemas mentales, la cognición ganaría en flexibilidad. Es importante tener presente que: “los módulos conceptuales

podrían tomar información de más fuentes y controlar más de un simple procedimiento motor, pero aún así seguirían siendo específicos de dominio, automáticos, rápidos, etcétera” (Sperber, 1994, p. 44).

Aunque la tesis de modularidad de los sistemas cognitivos centrales es independiente de las apelaciones adaptacionistas, en los últimos 20 años ambos argumentos han estado profundamente vinculados gracias a la “psicología evolutiva”; un programa de investigación que se autoproclama como la “nueva síntesis” al poner el contacto los desarrollos surgidos de la teoría computacional clásica con la tradición adaptacionista de la biología evolutiva (Buss, 2008; Cosmides y Tooby, 1994; 1997; Pinker, 1997). En este sentido Cosmides y Tooby (1994) argumentan que, del mismo modo que el resto de las estructuras anatómicas, la mente/cerebro estuvo sujeta a una serie de presiones adaptativas que configuraron su arquitectura actual, por lo que si queremos comprender su funcionamiento, debemos tener presentes los problemas adaptativos que le dieron origen.

Hasta este punto, contamos con una arquitectura cognitiva compuesta, principalmente, por sistemas de entrada y módulos centrales. Ambos son entendidos como mecanismos computacionales específicos de dominio (*i.e.*, como operaciones formales sobre información sintácticamente estructurada). Sin embargo, los “módulos chomskianos” aún no han entrado juego como parte de la arquitectura cognitiva: es necesario cerrar la brecha entre estos mecanismos de procesamiento de información y los sistemas de creencias.

3.1.3.2. *Teorías implícitas y desarrollo cognitivo*

Dos conceptos han estado dando vueltas a lo largo de mis reflexiones sobre las propuestas funcionalistas de la arquitectura mental, el de “módulo” y el de “especificidad de dominio”. Como ha de ser claro, la especificidad de dominio es una característica central de los módulos mentales y, en tanto propiedad, puede ser predicada también de otras entidades, entre ellas, las teorías científicas.

Recordemos que en capítulo 1 hablaba de la forma en la teorización en filosofía de la ciencia llegó a estar tan estrechamente vinculada a estudio de la mente que por momentos resulta simplemente imposible distinguir entre una y otro. Uno de los ejemplos más claros de la *epistemologización* de los fenómenos psicológicos es el estudio de las llamadas “teorías intuitivas” (Carey, 2009; Gopnik y Meltzoff, 1999; Gopnik y Wellman, 1994). Una cuestión muy interesante en esta área en el estudio del desarrollo cognitivo es que recibe apoyo tanto de AI como del AS a favor del modelo del lenguaje: (1) el AS se manifiesta en la idea de que las “teorías intuitivas” se organizan como los módulos computacionales, así como en el supuesto de que las teorías (científicas e intuitivas) son entidades cuasilingüísticas claramente especificables y (2) el AI está reflejado en la adopción de la metodología piagetiana al estudio de las concepciones infantiles y en la extrapolación de la visión historicista del desarrollo científico (Kuhn, 1970) al terreno del desarrollo cognitivo (véase Gopnik y Wellman, 1994).

Existen tres perspectivas principales en el estudio de la “teorías implícitas” (véase Pozo, 2003, capítulo 4, para una clasificación un tanto distinta):

1. *Las teorías implícitas son mecanismos computacionales:* Alan Leslie (1994) defiende que las “teorías implícitas” son, en sí mismas, mecanismos computacionales específicos de dominio. En este sentido Leslie (1994) comenta: “como resultado de la evolución adaptativa, el niño es un procesador especializado de información con una arquitectura que (en parte) refleja las propiedades del mundo” (p. 119). A grandes rasgos, existen tres módulos mentales que Leslie (1994) engloba bajo la noción de “Agencia”: ToBy (especializado en el procesamiento de las interacciones mecánicas entre los cuerpos); ToMM₁ y ToMM₂ (implicados, ambos, en el procesamiento de las interacciones intencionales entre agentes). Estos sistemas estarían implementados en la mente/cerebro al momento del nacimiento, aunque entrarían en juego en diferentes instantes de la ontogenia debido a aspectos de maduración y desarrollo cognitivo. El compromiso innatista de esta tradición nos remite inmediatamente a las propuestas de Fodor (1983) y, sobre todo, Chomsky (1957/2002; 1965): ToBy, ToMM₁ y ToMM₂ podrían bien ser entendidos como “módulos chomskianos”.
2. *La teorías implícitas son cuerpos de conocimiento autónomos; restringidos, pero no determinado, por el contenido mental previo a la experiencia:* Gopnik y Meltzoff (1999) sostienen que: “las estructuras cognitivas de los niños, como las de los científicos, son teorías;

que su desarrollo conceptual consiste en la formación y el cambio de teorías, y que su desarrollo semántico depende de una teoría” (p. 23) (véase también Gopnik y Wellman, 1994). De este modo, se entiende que en todos y cada uno de los niveles de la cognición — desde el aprendizaje y hasta el razonamiento— operan mecanismos análogos a los implicados en el surgimiento y cambio de las teorías científicas. Gopnik y Meltzoff (1999) y Gopnik y Wellman (1994) aceptan la posibilidad de los niños nazcan ya con *teorías* sobre el funcionamiento del mundo, sin embargo, estos autores no otorgan ningún estatus especial a estos cuerpos de conocimiento “innatos”, por lo que estarían sujetos exactamente a los mismos procesos de revisión y cambio que cualquier otra “teoría intuitiva” desplegada en un momento posterior del desarrollo cognitivo.

3. *La organización mental está compuesta tanto por “dominios centrales” como por “teorías intuitivas”*: Susan Carey (2009) y Carey y Johnson (2000), por su parte, defiende una postura a la mitad de las dos anteriores. Primero, sostienen que los niños nacen con sistemas conceptuales muy ricos y complejos organizados en torno a “dominios nucleares” (*core domains*) que tienen una larga historia evolutiva (*i.e.*, están presentes en otras especies) y permanecen relativamente constantes a lo largo del desarrollo. Por otra parte, existirían cuerpos de conocimiento construidos a partir de la interacción entre la “cognición central” (*core cognition*) y la experiencia que cada uno de nosotros tiene en el mundo. Para Carey (2009) y Carey

y Johnson (2000) el *cambio conceptual* (*i.e.*, el cambio teórico) es un proceso mayormente circunscrito al contexto de las “teorías implícitas”, donde puede haber casos de genuina “incomensurabilidad local” entre un sistema de creencias y otro como resultado del desarrollo de nuestra comprensión en un dominio de conocimiento. Por su parte, la “cognición central” suele ser más estable, estando sujeta más a un proceso de *enriquecimiento* (*i.e.*, a la adquisición de nuevas creencias formuladas a partir de un repertorio conceptual constante) que de genuino cambio conceptual.

Entonces, ¿cuántas y cuáles son las “teorías implícitas”? Resulta obvio que la respuesta a esta pregunta depende de la forma en que uno esté comprendiendo estos sistemas de conocimiento. A pesar de ello, todas las posturas coinciden en la necesidad del postular como “conocimientos centrales”: (1) un sistema de creencias sobre las interacciones mecánicas entre los cuerpos del *mesocosmos*, o “física intuitiva”, y (2) una “psicología intuitiva” (o, de forma más popular, “teoría de la mente”), es decir, un sistema de creencias sobre el comportamiento y las interacciones entre las personas. Además, hay quienes defienden también como conocimiento nuclear el conocimiento biológico (Keil, 1994; Mithen, 1998) —aunque Carey (2009) prefiere considerarlo como una diferenciación tardía del conocimiento social—, y el conocimiento de los números (Carey, 2009) —que Pozo (2003), por su parte, considera como un elemento del conocimiento físico. Con respecto al lenguaje, y de modo un tanto paradójico, se ha optado considerarlo como

un módulo periférico subsidiario del conocimiento social (Sperber, 1996; Tomasello, 2008).

Debería ser claro que aquí existe una confusión entre “teorías implícitas” en tanto *mecanismos de procesamiento de información*, o módulos, (Leslie, 1994; Carey y Johnson, 2000 en lo que tiene que ver con los “dominios nucleares”) y “teorías implícitas” como *sistemas de creencias* estructurados lógicamente (Gopnik y Meltzoff, 1999; Gopnik y Wellman, 1994). Ya en el trabajo pionero de Chomsky (1965) es evidente que la distinción entre *conocimiento del lenguaje* y *mecanismos computacionales para el procesamiento lingüístico* no siempre resulta nítida, situación que ahora se ha extendido a una gran parte de la investigación sobre dominio especificidad.

Así pues, en el estado actual de la investigación en ciencia cognitiva, persisten una serie de ambigüedades conceptuales que requieren ser aclaradas: ¿son los “módulos chomskianos” —entendidos como cuerpos de conocimiento— modulares en algún sentido interesante de la palabra?, ¿a cada “teoría intuitiva” le corresponde un módulo computacional?, ¿el desarrollo de una “teoría intuitiva” conlleva la aparición de nuevos mecanismos computacionales en la mente/cerebro?

Dar respuesta a estos cuestionamientos es algo que está más allá de los objetivos de este trabajo, por lo que me limitaré a presentar una alternativa que considero prometedora (aunque, como quedará claro en los capítulos siguientes, no comparto la opinión de que el conocimiento cotidiano se organice a partir de teorías). Ángeles Eraña (2006) sostiene que:

[Las] estructuras cognitivas que subyacen en los distintos tipos de proceso de desarrollo cognitivo (*i.e.*, enriquecimiento o cambio conceptual) son rígidas y se mantienen estables a través del desarrollo. En otras palabras, si bien puede haber distintos tipos de proceso que den cuenta de la adquisición y el desarrollo del conocimiento, esto no implica que necesariamente tengan que emerger estructuras cognitivas nuevas. Las primarias o básicas (los módulos computacionales) parecerían, al menos en principio, suficientes para explicar cómo estos diferentes tipos de proceso pueden desarrollarse (p. 97).

De este modo, a partir de una arquitectura cognitiva relativamente constante se puede muy bien dar cuenta de la adquisición y el desarrollo del conocimiento. No se trata aquí de negar que la experiencia juegue un papel muy importante en organización global de nuestra cognición. Si nos tomamos en serio que, aunque sin poder ser reducidos a ellos, los fenómenos cognitivos están fuertemente enraizados en procesos neurofisiológicos (supuesto que a estas alturas resulta hasta ocioso discutir), debemos de reconocer que, siendo el cerebro un sistema plástico, resultaría un sinsentido sostener que los procesos cognitivos sean rígidos (Carr, 2010). No obstante, plasticidad no significa omnipotencia, y el hecho de aceptar que el medio tiene profundas repercusiones sobre el modo en el que actuamos y pensamos no nos compromete con la idea de que la experiencia (por sí misma) hace surgir en nuestro cerebro nuevos mecanismos computacionales específicos de dominio.

Es importante dejar claro que la correspondencia entre mecanismos computacionales y sistemas de creencias no tiene porque ser 1:1. Es decir, sostener la existencia una “teoría implícita” sobre el funcionamiento del mundo social (una teoría de mente) no implica de ninguna forma defender que instaurado en la mente/cerebro deba de haber un módulo unitario encargado de procesar los estímulos sociales. Considero mucho más prometedora la idea de que nuestro conocimiento —esté o no estructurado a partir de teorías— es el resultado del funcionamiento coordinado de un conjunto limitado de mecanismos computacionales relativamente autónomos.

Resumiendo, emanada de la teoría computacional clásica tenemos una caracterización de la mente que podemos clasificar como lingüística en dos sentidos relacionados —pero no coextensivos— en la que los fenómenos cognitivos son vistos como: (1) procesos computacionales especificables en términos de un lenguaje formal (*i.e.*, módulos), y (2) como sistemas de conocimiento conformados por proposiciones lógicamente relacionadas (*i.e.*, teorías implícitas). En el capítulo siguiente exploro a detalle el impacto del AS sobre el que, considero, es el otro gran programa de investigación de los fenómenos cognitivos emanado del modelo del lenguaje: el análisis experimental de la conducta.

CAPÍTULO CUATRO

MODELOS MATEMÁTICOS EN EL ESTUDIO DE LOS PROCESOS COGNITIVOS: EL AS Y EL ANÁLISIS EXPERIMENTAL DE LA CONDUCTA

We do not check to see whether mathematical proof or laboratory investigation or statistical 'studies' are the right way to reason: they have become (after fierce struggles) what it is to reason rightly.

Ian Hacking

Al revisar el desarrollo de las concepciones modernas del lenguaje, la ciencia y la mente, salta a la vista un hecho que resulta, a todas luces, paradójico. El empirismo lógico defiende que los saberes científicos son los únicos que merecen ser llamados conocimiento. Esta serie de proposiciones empíricamente verificables reflejarían fielmente la estructura del mundo al estar vinculadas a través de la lógica y las matemáticas, por lo que se podría decir que la ciencia es, en gran medida, la *reconstrucción racional del mundo*.

Sin embargo, los positivistas sostienen al mismo tiempo un antipsicologismo exacerbado que los lleva a rechazar cualquier pretensión de que los fenómenos psicológicos deban de ser explicados a través de recursos conceptuales más allá de los que proporciona la descripción detallada de la conducta directamente manifiesta. En lo que tiene que ver con la psicología, los empiristas lógicos son, ante todo, *conductistas lógicos*.

De este modo, al querer dar cuenta del desarrollo del conocimiento científico de una forma naturalista (que, en última instancia, sería la única forma posible de hacerlo), nos hallamos ante el reto de explicar cómo es que un sistema de conocimiento lógicamente articulado, como se pretende es la ciencia, pudo ser desarrollado por organismos incapaces de representar el mundo internamente y carente de mecanismos cognitivos para la formulación de hipótesis y la deducción lógica. Es esta brecha entre el conocimiento y los organismos que conocen la que la teoría computacional clásica pretende subsanar.

La “revolución cognitiva” es heredera del empirismo lógico dado que, aceptando esta caracterización lingüística del conocimiento, se da a la tarea de desarrollar una teoría de la mente que encaje con esta visión formalista de la ciencia: si el conocimiento científico es una entidad cuasilingüística constituida por proposiciones sobre el mundo lógicamente relacionadas, entonces el resto de nuestros saberes deberían ajustarse a esta descripción. Dada esta caracterización logicista del conocimiento, resulta casi natural afirmar que los procesos psicológicos que sustentan su surgimiento y desarrollo son de naturaleza sintáctica. Los adelantos de la informática y la inteligencia artificial sentaron los cimientos sobre los que descansa el edificio conceptual de la ciencia cognitiva. Reconstruir esta historia fue el objetivo del capítulo anterior.

Sin embargo, el modelo del lenguaje no se agota con la caracterización de los agentes cognitivos. Al margen de todas estas discusiones sobre la arquitectura simbólica de la mente, la concepción formalista de la ciencia

continuó creciendo y fortaleciéndose —sobreviviendo, incluso, a las teorías historicistas y sociológicas sobre el conocimiento. En ciencias cognitivas, esta tradición intelectual encuentra una de sus expresiones más claras en el análisis experimental de la conducta.

Dicho de otro modo, el modelo del lenguaje tiene profundas repercusiones tanto sobre la caracterización del objeto de estudio de las ciencias cognitivas (los fenómenos mentales), como sobre las propiedades que se consideran deseables (y exigibles) de las descripciones naturalistas que se puedan hacer de él. En este capítulo, me concentraré en clarificar las implicaciones del segundo elemento de este binomio.

4.1. El contraataque empirista: el análisis experimental de la conducta

Si la revolución cognitiva de mediados del siglo XX es en gran medida una respuesta a la hegemonía del conductismo en el estudio de los fenómenos psicológicos, el análisis experimental de la conducta (en adelante, AEC) puede ser visto, a su vez, como la respuesta del conductismo a la revolución cognitiva.

Los analistas conductuales aceptan, al menos en términos generales, las críticas hechas por el funcionalismo computacional a las teorías asociacionistas clásicas. En lo específico, reconocen que es importante “abrir la caja negra”, aunque también insisten en que es preciso dotar de mucho mayor rigor axiomático a las teorías psicológicas (Luce, 1995; Myung y

Pitt, 2001). Cómo debe resultar evidente, para el AEC el comportamiento sigue siendo el objeto de estudio por excelencia para la psicología; sin embargo, ante la avasalladora evidencia que ha mostrado que los principios básicos del aprendizaje asociativo son insuficientes para explicar los datos empíricos, los analistas conductuales han considerado prudente incorporar las *restricciones* impuestas por la arquitectura cognitiva como un elemento importante en *algunos* de sus modelos teóricos.

Es importante dejar claro que el modelamiento matemático en el AEC va mucho más allá del uso de ecuaciones como simples herramientas teóricas. Más que un viraje metodológico, lo relevante del AEC es su cambio de actitud hacia el lenguaje psicológico (*i.e.*, en el uso de *términos mentalistas* como elementos explicativos de la conducta). Lo cierto es que hasta antes de la década de 1960 “lo psicológico” era algo con lo que se tenía que lidiar en el estudio científico del comportamiento, mientras que a partir de esos años, con el nacimiento de las ciencias cognitivas, las explicaciones de la conducta pasaron a ser *orgullosamente psicológicas*. La propuesta de Turing (1950) de tratar a los fenómenos psicológicos como procesos computacionales dotó al estudio de la cognición de la respetabilidad científica que tanto le hacía falta. En este sentido, las explicaciones ofrecidas por el AEC suelen ser explicaciones genuinamente cognitivas.

La consolidación del AEC no es algo que se pueda ubicar en un momento específico. Desde la misma década de los 60s, comenzó un proceso de reconfiguración profunda al interior de las tradiciones conductistas — hecho que no es para nada menor, ya que echa por la borda la idea simplista

de que la “revolución cognitiva” representó un *cambio de paradigma* en el estudio de los fenómenos psicológicos. No obstante, si uno tuviera que elegir un evento para “conmemorar” el nacimiento del AEC “moderno”, éste bien podría ser la publicación del tan celebrado trabajo de Robert A. Rescorla y Allan R. Wagner (1972) sobre los mecanismos implicados en el aprendizaje por condicionamiento clásico. Rescorla y Wagner (1972) incorporan en una simple expresión (con toda la formalidad del modelamiento matemático) algunas de las preocupaciones cognitivas que había comenzado a adquirir relevancia en el estudio del aprendizaje desde mediados del siglo XX.

A diferencia de las aproximaciones formalistas previas —algunas de las que habían gozado de muy buen éxito, como la ley de la igualación de Herrnstein (1961, citado en Jozefowicz y Staddon, 2008), que describe la distribución que un organismo hace de su conducta entre diferentes actividades como una función de la tasa de reforzamiento que le ofrece cada una de ellas (*i.e.*, como consecuencia de factores totalmente externos)—, en el modelo de Rescorla y Wagner (1972) se entiende al aprendizaje como un resultado de la interacción entre la estructura del ambiente y *factores cognitivos*, intrínsecos a la arquitectura mental del organismo.

Esto es de vital importancia dado que, para que una expresión matemática forme parte de una genuina explicación psicológica, al menos algunos de sus parámetros deben representar *variables genuinamente psicológicas*. De este modo, en el modelo de Rescorla y Wagner (1972) establece que:

$$\Delta V = k (\lambda - V)$$

Dónde el valor “k” representa una constante resultada de la notoriedad del EC y el EI (*i.e.*, $k = \alpha ECs \cdot \beta EI$); “V” representa el valor asociativo de un EC (o ECs) resultado de la experiencia previa de organismo con este estímulo; “ λ ” constituye el valor efectivo del EI en el ensayo particular que está teniendo lugar; y “ ΔV ” es el cambio en el valor asociativo del EC (o ECs) producto de ese ensayo específico.

Es importante hacer notar que el valor en el parámetro “V” (o su cambio) no es un elemento inherente al estímulo mismo, sino algo *que el organismo le atribuye* como resultado de su experiencia con él. *Teóricamente* se postula que el parámetro “V” representa las *expectativas que el organismo sostiene* con respecto a la ocurrencia de un determinado evento dada una determinada configuración ambiental, y que el resultado de “ $\lambda - V$ ” (la diferencia entre lo que ocurre y lo que se esperaba) es un indicador de lo “sorprendente” que resulta un evento *para el organismo*. De este modo, se entiende que entre mayor sea el grado de sorpresa de un hecho, más es lo que se puede aprender de él. Cuando las expectativas del organismo predicen casi a la perfección lo que va a ocurrir ($\lambda \approx V$) entonces el nivel de sorpresa es mínimo ($\lambda - V \approx 0$) y no tiene lugar aprendizaje alguno. La constante “k” no deja de ser un parámetro importante desde el punto de vista psicológico, ya que incorpora la sensibilidad del organismo para percibir el ambiente como algo que posibilita todo el proceso de aprendizaje. Si “ $k \approx 0$ ” (por ejemplo, si el evento se sitúa más allá del espectro perceptual accesible para organismo) automáticamente “ $\Delta V \approx 0$ ” y el organismo no extrae información alguna del evento.

Es importante dejar claro que el cambio de actitud del AEC hacia los mecanismos psicológicos (en tanto procesos computacionales) no conlleva el abandono de los modelos y las propuestas teóricas previas. La actitud de los analistas experimentales al respecto es más bien pluralista, una tolerancia teórica que en el discurso promueve el trabajo transdisciplinar. A diferencia de las teorías cognitivas clásicas, para el AEC, la representación y la computación “entra ahí donde es necesaria” con fines explicativos, no como un presupuesto conceptual previo a la formulación de cualquier modelo teórico.

De este modo, al igual que las teorías computacionales del funcionamiento psicológico, el AEC recibe con los brazos abiertos los desarrollos surgidos de otras disciplinas como la inteligencia artificial, la neurofisiología, la biología evolutiva y, obviamente, las matemáticas; pero siempre insistiendo en que el comportamiento es siempre su objeto último de estudio y en la importancia de que las teorías psicológicas se presenten de la manera más formal posible. Como comentan Myung y Pitt, 2001:

Los modelos matemáticos, que buscan descripciones cuantitativas de los datos, intentan caracterizar patrones de comportamiento explorando directamente la forma de los mecanismos subyacentes que le dan lugar a la conducta de interés. Esto incluye cuestionamientos sobre cómo la información de los estímulos está siendo representada, qué computaciones son llevadas a cabo sobre los *inputs*, qué mecanismos están implicados en el procesamiento de la información, y otras cuestiones por el estilo (p. 430).

Es interesante hacer notar los paralelos existentes entre el desarrollo de las concepciones modelísticas en filosofía de la ciencia y el surgimiento la consolidación del uso de modelos matemáticos por parte del AEC (véase el capítulo 6 de Moulines, 2011, sobre la llamada “concepción semántica” en filosofía de la ciencia). En ambos contextos, una tradición intelectual directamente inspirada en la visión positivista del mundo y de la ciencia se ven obligadas a incorporar los desarrollos teóricos surgidos de posiciones incluso antagónicas. En ambos casos, se da pie a un pluralismo teórico que enmascara un compromiso formalista aún más fuerte que el del positivismo lógico.

Sin embargo, hay una diferencia esencial que no puede ser pasada por alto: mientras que en la concepción semántica puede observarse una tendencia hacia un pragmatismo antirrealista cada vez más fuerte (véase, por ejemplo, van Fraassen, 1977), en el AEC persiste, muchas veces de forma implícita, cierto realismo ingenuo, no sólo con respecto a las teorías científicas, sino también los parámetros postulados por cada modelo.

4.2. Compromisos ontológicos en el AEC: ¿son los modelos matemáticos sustitutos de las descripciones verbales?

Las similitudes entre las tradición modelística de van Fraassen (1977) en filosofía de la ciencia y el AEC son más transparentes si consideramos que es la economía —la disciplina social más estrechamente ligada a las reflexiones epistemológicas sobre el uso de modelos— de donde los analistas de la

conducta retoman las ideas clave sobre las que justifican su aproximación al estudio de la conducta. No es gratuito que Howard Rachlin (1995), uno de los investigadores más representativos del AEC, sea también reconocido como uno de los pioneros en el campo de la *economía conductual*. Dados los paralelos entre uno y otro campo de conocimiento, una revisión de las consideraciones conceptuales que están teniendo lugar en economía nos permitirá formarnos una idea más clara de los retos a los que se enfrenta el AEC cuando defiende el uso de modelos matemáticos como un elemento clave en la investigación de los fenómenos psicológicos.

En el desarrollo histórico de la economía, el siglo XIX estuvo marcado por la transición de las descripciones verbales de los fenómenos a los modelos matemáticos (Morgan, 2009). En psicología, por su parte, persiste una fuerte disputa por determinar cuál es la mejor forma para describir los fenómenos cognitivos, particularmente en lo que tiene que ver con el uso (y abuso) de una psicología planteada en términos de deseo-creencia (lo que típicamente se conoce en filosofía como una “psicología popular”). Esta idea queda claramente reflejada en los comentarios de Wasserman y Zentall (2009) sobre las diferencias entre los campos de estudio de la *etología cognitiva* y la *cognición comparativa*.

El rasgo distintivo de la etología cognitiva (Griffin, 1994; de Waal, 2006) es la adscripción de habilidades y procesos psicológicos típicos de los humanos a otros animales. De ese modo, de Waal (2006) sostiene que los chimpancés realizan juicios morales análogos a los que tienen lugar en la especie humana. Wasserman y Zentall (2009) critican este tipo de pos-

turas por considerar que se basan casi exclusivamente en anécdotas imposibles de reproducirse en ambientes experimentales controlados y que, aún en los casos en los que se trate de fenómenos bien documentados, no es necesaria toda la parafernalia del lenguaje mentalista para dar cuenta de ellos:

El estudio experimental de la inteligencia animal deberá iluminar enormemente nuestra comprensión de las adaptaciones conductuales y su evolución en el reino animal. Quizá deberíamos concentrar nuestros esfuerzos en esta tarea y dejar las especulaciones mentalistas a los filósofos, cuyas teorías de la mente y conjeturas sobre la conciencia no necesitan estar limitadas por las restricciones de la ciencia natural (p. 10).

¿Qué hacer, entonces, si nuestras descripciones cotidianas de la conducta son inadecuadas para el estudio científico de los fenómenos cognitivos? Las opciones ante esta cuestionamiento van, como ya se comentó, desde aquellos que proponen que tratar los estados mentales como estados funcionales nos permite desarrollar una psicología naturalista que dé cuenta de la conducta en términos de deseo-creencia (Fodor, 1983, Putnam, 1963/1983), hasta aquellos que sostienen que hay que abandonar definitivamente la “psicología popular” y sustituirla por las descripciones genuinamente materialistas de la neurociencia (Churchland, 1981).

La postura del AEC puede ser ubicada a la mitad entre estos dos extremos. No se trata de eliminar los términos mentales de las explicaciones de la conducta, sino de tratarlos como constructos teóricos sólidos, con el rigor propio de la ciencia natural. Así pues, los peligros para una ciencia

del comportamiento no están tanto en los términos que emplea, sino en *cómo* los emplea. Tomando a la economía como ejemplo, los analistas de la conducta sostienen que: “la falta de precisión es un defecto serio del modelamiento verbal” (Myung y Pitt, 2001). De este modo, mientras que las descripciones verbales de los fenómenos serían adecuadas en las primeras etapas de investigación de un fenómeno, una vez que hemos ganado un conocimiento suficiente sobre su estructura, resulta más provechoso echar mano de las herramientas formales de la matemática:

Dado que el objetivo del modelamiento matemático es especificar los detalles en la caja negra, no resulta la mejor estrategia en las fases tempranas y exploratorias de la investigación, cuando un fenómeno es explorado por primera vez. Por el contrario, el modelamiento matemático es probablemente más fructífero en las etapas más avanzadas de la investigación, cuando ha sido adquirido un conocimiento considerable sobre la dinámica de la conducta a través del modelamiento verbal (p. 430).

Tal cual está planteado, el razonamiento a favor del modelamiento matemático desarrollado por los analistas de la conducta resulta muy atractivo. No obstante, es importante no confundir una justificación con una práctica científica. Un primer elemento a tener en cuenta es que, como comenta Mary S. Morgan (2009), la transición entre las descripciones verbales y los modelos matemáticos no es tan natural como se pretende. La adopción de los modelos matemáticos en economía significó un cambio radical en el *estilo de razonamiento* que tiene que ser analizado en términos históricos:

Una vez aceptado por un grupo de científicos, un estilo de razonamiento termina por ser natural para ellos, tan natural que ya no lo cuestionan. No cuestionan ni sus orígenes históricos, ni la objetividad del conocimiento ganado al usarlo como método, y tampoco apelan a ningún nivel superior o externo para su justificación. Esto es porque, como Hacking argumenta, una vez que un estilo de razonamiento es aceptado en una comunidad, razonar correctamente significa razonar en ese estilo (p. 17).

Esto es así hasta tal punto que la comprensión de un determinado modelo económico está en gran medida condicionada a al conocimiento del entorno (no sólo teórico, sino también histórico y social) en que el éste surgió. De esta forma, existen registros históricos de modelos cuya aplicación permanece como un misterio a falta de información suficiente que nos permita reconstruir el contexto que le dio origen (Morgan, 2009). Entender que el mundo y la ciencia son cosas distintas es de vital importancia para impulsar un desarrollo real del conocimiento científico. Tenemos que aceptar que, si en verdad es el caso que dios escribió la naturaleza en matemáticas, somos lectores bastante incompetentes.

La cuestión del estatus ontológico de los modelos científicos no es para nada un asunto menor. Y es que, como hace ver Morgan (2009), si un modelo es una representación adecuada de la realidad (en algún sentido relevante), *entonces* los resultados obtenidos a través de su aplicación podrían ser extendidos de forma directa al mundo; no obstante, defender la tesis detrás de este “si” representa un compromiso muy fuerte. La historia

de la ciencia nos ha enseñado una y otra vez que “el libro de la naturaleza” —otra metáfora lingüística, por cierto— es mucho más complejo e interesante de lo que pensábamos (Olson, 1994/1998). Tomar como verdadero y acabado un determinado modelo o teoría científica es el camino más corto que se puede tomar para transitar del optimismo a la frustración intelectual.

En un contexto con implicaciones epistemológicas aún más profundas, Nancy Cartwright (1980/1991) afirma que es un error considerar que las reconstrucciones causales en términos verbales pueden ser simplemente sustituidas por *leyes generales* (entendidas como expresiones matemáticas que pretenden subsumir la estructura de los fenómenos naturales), ya que las leyes, a diferencia de las reconstrucciones causales, pueden ser *redundantes*: diferentes leyes pueden igualmente bien dar cuenta de un determinado fenómeno. Dado que sólo se puede inferir la verdad de una explicación en la medida en que no haya una alternativa que describa igualmente bien el hecho bajo estudio, predicar la verdad de las leyes aisladas falla al no tomar seriamente cómo tiene lugar la práctica explicativa en ciencia.

La “competencia” entre abordajes teóricos —aproximaciones que suscriben leyes diferentes para el mismo fenómeno— es fomentada en la física, pero sólo es permitida una historia causal simple. Aunque los filósofos generalmente creen en las leyes y niegan las causas, la práctica explicativa en la física es exactamente la contraria (Cartwright, 1980/1991, p. 386).

Si esto es verdad para las leyes en física, tanto más se puede decir de disciplinas como la economía o las ciencias cognitivas, cuyos modelos resultan idealizaciones extremas de procesos influenciados por múltiples variables que, además, apenas estamos comenzando a comprender. Se puede objetar que las leyes y los modelos matemáticos no son elementos necesariamente coextensivos; sin embargo, a este respecto y tal cual está planteada la crítica de Cartwright (1980/1991) —a decir, que las leyes de la física no contienen en sí mismas información sobre la estructura de la naturaleza—, no encuentro una sola razón para suponer que esto haya de ser diferente en el caso de los modelos matemáticos en economía o ciencia cognitiva. La sola distinción conceptual entre modelos matemáticos y leyes generales es algo que merece mucha mayor atención por parte de los analistas conductuales, quienes suelen emplear ambos términos de manera indistinta.

De este modo, estoy de acuerdo con Myung y Pitt (2001) en que, para desarrollar un modelo matemático que ilumine nuestra comprensión sobre un determinado proceso cognitivo, es necesario haber ganado conocimiento suficiente sobre él a partir de medios “más informales”. No obstante, no comparto el supuesto de que, una vez que se tiene una expresión matemática que “describe” un determinado hecho, se pueda prescindir del resto del conocimiento que de él se posee. No puedo más que reconocer que las matemáticas y la lógica son herramientas muy útiles en nuestros esfuerzos por comprender el mundo, pero es muy importante nunca perder de vista que *hacer ciencia no es hacer matemáticas o lógica*.

Un ejemplo muy claro de que los modelos formales desarrollados en ciencia pueden ser comprendidos (y empleados) con relativa independencia de los supuestos causales que, se podría pensar, los constriñen y determinan los criterios de su aplicación, es la descripción que hace José Luis Bermúdez (2003) de la “teoría de forrajeo óptimo”, tan celebrada por los analistas del comportamiento:

La evolución ha trabajado de tal manera que (al menos para los proponentes de la teoría del forrajeo óptimo) las especies forrajeadoras han evolucionado conjuntos de estrategias heurísticas que resultan en una adaptación óptima a sus nichos ecológicos. Esta adaptación óptima puede ser modelada matemáticamente en términos de lo que es, en última instancia, una versión sofisticada de la teoría de la utilidad esperada, pero las conductas en que ésta se manifiesta no resultan de la aplicación de tal teoría (p. 115).

El hecho de que a los patrones conductuales mediados por heurísticos puedan ser modelados a partir de los mismos supuestos de la teoría de la utilidad esperada es sumamente revelador si tenemos en consideración que, como argumenta Gerd Gigerenzer (2004; Gigerenzer y Brighton, 2008; Gigerenzer y Gaissmaier, 2011; Todd y Gigerenzer, 2000), las explicaciones del comportamiento en términos de heurísticos (o de *racionalidad limitada*) pretender ser la *antítesis* de la teoría de la maximización. Aún más dramático resulta el cada vez más abundante cuerpo de evidencia convergente que nos muestra que la idea de un agente totalmente racional detrás de la noción de *homo economicus* —el pilar fundamental sobre el que descansa

buena parte de la teorización desarrollada en economía y en ciencia cognitiva— ¡bien podría no ser más que una ilusión! (Gigerenzer, 2004; Damasio, 1994/2006; Tverky y Kahneman, 1974, Kahneman y Tverky, 1984; Sutherland, 1996; y los trabajos en Bargh, 2007; Hassin, Uleman y Bargh, 2005).

El hecho de que un modelo pueda gozar de un relativo éxito teórico y predictivo al margen de su grado de contacto con la realidad que pretende describir, tiene que ser entendido como un aspecto estructural de la práctica científica. Esto sólo cobra sentido si reconocemos que, siguiendo a Morgan (2009), el ser un “buen investigador” no sólo recae en el rigor de individuos particulares para aproximarse experimentalmente a la naturaleza, recabar datos y “acomodarlos” en teorías que reflejen fielmente la estructura del mundo. Por el contrario, debemos tener presente que la ciencia es esencialmente una actividad creativa (Morgan, 2009) y socialmente distribuida (Harris, 1994; Eraña y Martínez, 2004).

Tan importante como su aspecto empírico, la labor científica y filosófica depende de la capacidad de los investigadores de encontrar analogías entre el hecho de su interés y las descripciones surgidas en otros campos que, a primera vista, podría parecer que no guardan relación alguna con aquello que se pretende explorar (Hesse, 1966, citado en Morgan, 2009). El “modelo del lenguaje” es particularmente ilustrativo en este sentido: al tratar a la ciencia y a la mente como si estuvieran organizados como un lenguaje (algunas veces de forma directa, otras haciendo uso de una metáfora lingüística adicional, como el ordenador o el libro), filósofos y científicos

lograron iluminar nuestra comprensión en estos importantes campos del saber; aunque, como he venido sosteniendo, el propio desarrollo de nuestro conocimiento sobre un fenómeno muchas veces termina orillándonos a abandonar la metáfora original (posiblemente a favor de otra analogía más fructífera).

¿Qué papel juegan, pues, los modelos en la ciencia? En este sentido, Mary Morgan (2009) sostiene que más allá de ser *transcripciones* de la realidad, éstos se constituyen como *realidades* en sí mismos. El elemento esencial de los modelos (sean matemáticos o de cualquier otro tipo) es que son *manipulables*, es decir, que *se puede experimentar directamente sobre ellos*. Hay un aspecto que resulta muy interesante de esta propiedad:

Mientras que los experimentos que se llevan a cabo sobre un modelo pueden *sorprender* a los economistas con resultados inesperados, los experimentos de laboratorio pueden *confundir* al economista científico al producir resultados que no sólo son inesperados, sino también potencialmente inexplicables a partir del conocimiento existente (Morgan, 2009, p. 36).

De este modo, al explorar la forma en que un determinado modelo se comporta ante la variación sistemática de los valores de sus parámetros —un ejercicio que tiene lugar al margen de cualquier restricción impuesta por el mundo externo al modelo—, los científicos se pueden topar con predicciones contraintuitivas que pueden ser sometidas a experimentación empírica. Este hecho nos muestra claramente la relación entre realidad y teoría es

mucho más compleja de lo que originalmente se pensó en la filosofía de la ciencia clásica: el conocimiento científico no es un cúmulo de datos neutrales y lógicamente vinculados sobre la estructura de la naturaleza, sino que nuestras teorías y modelos influyen de manera profunda (¿determinan?) nuestra visión de los fenómenos y la forma en que encaramos su estudio.

Lo anterior lleva a Morgan (2009) a afirmar que nuestros modelos tienen una *doble vida*: son herramientas que los científicos usan para investigar la realidad, pero también —y sobre todo—, los modelos son objetos de investigación científica por derecho propio. Cuando un investigador explora la dinámica de sus modelos de la realidad, no gana conocimiento sobre la parcela de la naturaleza que éstos pretenden representar. Por el contrario, al proceder así lo que se intenta comprender son *las implicaciones de sus ideas sobre el mundo*. Gran parte de la actividad científica —acaso la más interesante y productiva— no pasa por la experimentación y la medición directa de la naturaleza, sino que es más bien un intento de comprender nuestros modelos y teorías (no al mundo, sino lo que del él pensamos) y, en última instancia, de comprendernos a nosotros mismos (nuestros intereses y curiosidades, nuestros anhelos y objetivos).

Como Jesse J. Prinz (2008, p. 205) lo dijera: “los datos sin teoría están vacíos y la teoría sin datos está ciega”. La estructura del mundo restringe las teorías que de él podemos formar pero nuestras teorías y modelos guían, a su vez, la forma en que comprendemos y nos aproximamos al estudio del mundo. De algún modo, los *marcos conceptuales* que posibilitan nuestras interacciones sociales —de las cuales la ciencia es sólo una, aunque una

particularmente interesante— nos permiten “recortar” la realidad de tal forma que nos sea posible referirnos a ella. Esta es la idea detrás del *realismo interno* postulado por Hilary Putnam (1990) y desarrollado por autores como León Olivé (1996), quien afirma que:

La expresión metafórica de que un marco conceptual “recorta” a la realidad de cierta manera, quiere decir que la realidad no está dividida en sí misma, en hechos y objetos que —como dice Putnam— sean auto-identificantes, con plena independencia de los marcos conceptuales. La realidad por supuesto está constituida por hechos y por objetos, pero no hay una única —y correcta— respuesta a la pregunta, ¿cuáles son los hechos y los objetos que constituyen la realidad?, sino que la respuesta depende del marco conceptual que se esté usando (p. 65).

Desmarañar los argumentos detrás de realismo interno es una tarea compleja que está más allá de los objetivos de este trabajo. No obstante, la idea de que la ciencia puede ser definida, esencialmente, como un proceso de comunicación racional (Olivé, 1996) resultará de suma importancia en los capítulos siguientes.

Por el momento, sólo me gustaría enfatizar que defender la existencia de una realidad externa e independiente, que restringen en gran medida los sistemas conceptuales que de ella podemos formar, no implica rechazar que nuestras teorías y modelos (y en general, nuestro conocimiento) influyan de manera profunda en cómo la comprendemos y exploramos. No podemos confundir la naturaleza con lo que podemos decir de ella. Las propuestas de Cartwright (1980/1991), Olivé (1996) y Morgan (2009) nos

hacen ver que seguir sosteniendo que el éxito de nuestras teorías recae en su *correspondencia* con el mundo resulta una sobresimplificación, tanto de la realidad —al considerar que nuestras idealizaciones científicas agotan su descripción—, como de la práctica científica —al sostener que ésta se lleva a cabo sólo a partir de hechos “comprobados”.

4.3. MODELAMIENTO MATEMÁTICO Y ALGORITMOS COMPUTACIONALES

En lo que va de este capítulo, he tratado a los modelos matemáticos como herramientas para la descripción de los fenómenos psicológicos que, en última instancia, son ajenos al objeto que describen. Es decir, los argumentos esgrimidos por los analistas de la conducta apelan, de forma típica, a la importancia de que nuestras teorías sobre la naturaleza sean formuladas en los términos más claros (*i.e.*, formales) posibles. Las ciencias cognitivas —como en su momento la física y la química— pueden beneficiarse mucho de la matematización de sus teorías, ya que esto permitiría procesos de comunicación mucho más eficientes y predicciones teóricas aún más precisas.

Sin embargo, en el contexto de la investigación psicológica, las expresiones matemáticas también pueden figurar *como parte de los fenómenos cognitivos*: siendo consideradas como *algoritmos* dentro de las teorías computacionales del funcionamiento mental. Esta aproximación de remonta al trabajo pionero de David Marr (1982/2010) sobre el procesamiento de la información visual. En su obra clásica, *Vision: A computational investigation into the human representation and processing of visual information*,

Marr (1982/2010) defiende la idea de que una explicación computacional completa debería estar formulada a partir de tres niveles explicativos interrelacionados: el computacional, el algorítmico y el físico:

En un extremo, el nivel superior, está la teoría computacional abstracta del dispositivo, en la que su desempeño es caracterizado como un mapeo de un tipo de información a otro, las propiedades abstractas de este mapeo están definidas de forma precisa, y su adecuación para la tarea en cuestión es demostrada. En el centro está la elección de las representaciones para el *input* y el *output*, así como el algoritmo que será usado para transformarlos uno en otro. Y, en el otro extremo, se encuentran los detalles de cómo el algoritmo y la representación están realizados físicamente —la arquitectura detallada de la computadora, por decirlo de algún modo. [...] Cada uno de estos tres niveles de descripción tiene su lugar en la eventual comprensión del procesamiento de la información perceptual, y por supuesto que los tres están tanto lógicamente como causalmente relacionados (p. 25).

El caso de David Marr es bastante atípico dentro las ciencias cognitivas. Su propuesta teórica podría muy bien ser considerada como el intento más serio y mejor logrado de poner el contacto las herramientas conceptuales de la filosofía de la mente con los desarrollos de la inteligencia artificial y la neurofisiología. Marr es un teórico que no sólo entiende a la perfección las implicaciones empíricas y conceptuales de cada uno de estos campos del conocimiento por separado, sino que muestra una gran claridad intelectual (que no puede menos que ser calificada como sorprendente) para poner el contacto los hallazgos surgidos de estas tres áreas e integrarlos en una

explicación coherente de la percepción visual. Lamentablemente, con su muerte a los 35 años de edad, la ciencia perdió de forma prematura a una de las mentes lúcidas del siglo XX.

La propuesta de Marr (1983/2010) es revolucionaria tanto en el contexto de la teoría computacional de la mente como en las neurociencias. De entrada, rechaza la idea —defendida por autores tan influyentes como Fodor (1975) y Putnam (1963/1983)— de que los fenómenos mentales puedan ser enteramente descritos en términos de relaciones entre estados funcionales y que el sustrato físico de la cognición, el “hardware”, es un elemento secundario a la explicación psicológica. Para él, el *nivel físico* (las estructuras cerebrales y los patrones de interconexión neural, en el caso de los sistemas biológicos) complementa al *nivel computacional* (el cómo y porqué el sistema hace lo que hace) y el *nivel algorítmico* (cómo es tiene lugar el procesamiento de información, qué tipos de representaciones entran en juego y qué procesos manipulan estas representaciones).

No hay nada que tenga que decir en contra de la propuesta de Marr (1983/2010); por el contrario, considero que ésta representa un camino que hoy, a 30 años de haber sido formulado, sigue presentándose como uno de los más prometedores para progreso de las ciencias cognitivas en su conjunto. Sin embargo, sí considero que su aplicación al estudio de fenómenos cognitivos concretos no siempre ha sido la más afortunada.

Un primer obstáculo tiene que ver con lo sofisticada que resulta la propuesta en sí misma; son pocos los investigadores que cuentan con los conocimientos conceptuales y metodológicos necesarios para llevarla a la

práctica. Por otra parte, empero todo el discurso (añejo, por cierto) sobre la importancia del trabajo transdisciplinar en ciencias cognitivas, la realidad nos muestra que en este sentido a penas vamos empezando y que, a pesar interesantes proyectos de genuina colaboración, aún son necesarios esfuerzos mucho más serios de integración.

Como lo comentan Penn y Povinelli (2007), en la tradición heredera de la teoría computacional clásica (que describí a grandes rasgos en el capítulo anterior) son más bien escasos los intentos de echar mano de métodos formales (como el modelamiento matemático) para la formulación y evaluación de sus teorías (aunque autores como Gopnik y Tenenbaum (2007) han avanzado interesantes propuestas sobre el uso de redes bayesianas para describir la adquisición y desarrollo del conocimiento infantil).

En términos generales, se podría decir que los expertos en psicología cognitiva tienen conocimientos neurofisiológicos muy básicos y que, de forma paralela, el grueso de los neurocientíficos carece una comprensión sólida en lo que tiene que ver con las teorías simbólico-representacionales del funcionamiento mental. Finalmente, la formación que tanto psicólogos como neurocientíficos poseen en lo que tiene que ver con el uso de sistemas formales de razonamiento (como lo son la lógica y las matemáticas) es, por decir lo menos, bastante pobre (algo que debería de hacernos reflexionar sobre la calidad de la instrucción que están recibiendo los futuros investigadores, aunque eso es, como se dice coloquialmente, “arena de otro costal”). Recordemos que Morgan (2009) insiste que la adopción generalizada de un sistema formal de representación no es un cambio metodológico menor,

sino que constituye una modificación radical en el *estilo de razonamiento* empleado para pensar y explorar el mundo. Podría decirse, pues, que las condiciones no son las idóneas para el aprovechamiento de un sistema explicativo tan sofisticado como el que propone Marr (1983/2010), aunque hay ciertos indicadores que nos permiten ser optimistas hacia el futuro.

Con todo y estas limitantes, sería un error afirmar que no ha habido esfuerzos por aplicar el esquema de Marr (1983/2010) al estudio y la explicación de fenómenos cognitivos concretos. Como habría de ser evidente, aquellos mejor capacitado para afrontar este reto (particularmente en lo que tiene que ver con el nivel *algorítmico*) son los analistas experimentales de la conducta, gracias a su interés y relativo dominio de las herramientas propias del modelamiento matemático. A pesar de ello, sostengo que el alcance de todos estos esfuerzos se ve bastante limitado en virtud de que se suelen pasar por alto una distinción conceptual muy básica que, sin embargo, es de vital importancia para el progreso de nuestra comprensión de la mente: un modelo del un proceso natural *es algo totalmente distinto* de un algoritmo cognitivo.

A primera vista, esto podría parecer un matiz trivial; no obstante, mostraré que un análisis un poco más detallado nos muestra que conlleva implicaciones teóricas muy profundas. A partir de la discusión sobre la naturaleza e importancia de los modelos científicos desarrollada en la sección anterior, debería de quedar claro que éstos son *herramientas epistemológicas* que, en conjunción con otros elementos de la práctica científica, soportan y enmarcan nuestros esfuerzos por comprender el mundo. Como Eraña y

Martínez (2004) lo hacen ver, el conocimiento humano va mucho más allá de un conjunto de proposiciones sobre la naturaleza. Más correcto sería decir que éste es un fenómeno que se distribuye en el entorno normativo en donde tienen lugar las interacciones entre personas, del cual los cerebros son sólo una parte y que incluye, de forma muy importante, los sistemas simbólicos y los artefactos desarrollados por la cultura (Clark, 1997, Eraña y Martínez, 2004). Aunado a lo anterior, la función y utilidad de un modelo científico trasciende ampliamente su grado de correspondencia con la realidad (Morgan, 2009).

El estatus ontológico predicable de un algoritmo cognitivo no podría ser más distinto. De la propuesta de Marr (1983/2010), se puede extraer que éstos son *reglas para el procesamiento de información* implementadas (no necesariamente de forma consciente) en el funcionamiento de la mente/cerebro. Dicho de forma más clara, a diferencia de los modelos, los algoritmos cognitivos *no son representaciones de procesos psicológicos*, sino que *forman parte de los procesos psicológicos* (*i.e.*, son susceptibles de ser representados y figurar *como elementos en modelos* sobre el funcionamiento mental).

En tanto partes constitutivas de los procesos psicológicos, defender la existencia de algoritmos cognitivos presupone un compromiso realista mucho más fuerte que el propio de los modelos. Esto podría parecer una contradicción con respecto a la anterior observación de que nuestros marcos conceptuales nos permiten “recortar” el mundo de diferentes maneras (*i.e.*, que nuestra *percepción* de la realidad está profundamente influenciada por

nuestras *ideas* sobre ella) (Olivé, 1996). No obstante, establecer una distinción entre *objeto de estudio* y *descripciones formales* en psicología no nos compromete con la idea de que los algoritmos cognitivos son entidades independientes de nuestros marcos conceptuales: tanto uno como las otras están *parcialmente determinados* por las concepciones que sostenemos cuando nos aproximamos al estudio de la naturaleza pero, aún así, su estatus ontológico es diferente.

Las ciencias cognitivas son particularmente vulnerables de perder esta distinción entre fenómeno y descripción. Por ejemplo, podemos tener una explicación que dé cuenta de la dinámica de los movimientos planetarios a partir de unas cuantas leyes generales (las leyes fundamentales de la física clásica, si queremos). Estas leyes estarían enmarcadas en un contexto epistemológico más amplio que determinaría los criterios para su aplicación y la evaluación de su éxito predictivo —recordemos que Cartwright (1980/1991) argumenta convincentemente que las leyes no contienen en sí mismas información sobre la estructura de la naturaleza, sino que forman parte de un entramado de práctica explicativas mucho más extenso. Pero nadie en su sano juicio estaría dispuesto a defender la idea de que los planetas “conocen-aplican-computan” las expresiones formales que describen sus patrones de movimiento. Un modelo de la dinámica planetaria (sea matemático o sea de cualquier otro tipo) no da cuenta de propiedades que sean predicables de los planetas individuales, sino que detalla las pautas de interacción entre diferentes cuerpos en función de sus masas y de la fuerza gravitacional.

Sin embargo, desde el momento en que la descripción de los fenómenos mentales empezó a depender básicamente de las mismas herramientas conceptuales empleadas para formular el conocimiento científico, distinguir entre fenómeno y teoría se ha vuelto una tarea bastante complicada. Las razones que justifican este supuesto han sido el hilo conductor para el desarrollo de este trabajo. Básicamente, tanto la ciencia como la mente han sido comprendidas como entidades lingüísticas, especificables en términos sintácticos.

Librar este obstáculo no es para nada una tarea sencilla. Aunque los algoritmos cognitivos sean, en sí mismos, parte de los fenómenos psicológicos, esto no quiere decir que tengamos acceso privilegiado y directo a ellos. A trepanar un organismo o ver sus patrones de activación neural en un resonador no vemos operaciones matemáticas resolviéndose a medida que se hace frente a una tarea; la introspección tampoco nos ayuda en este sentido: la mayor parte de nuestros procesos cognitivos están más allá de cualquier control o acceso consciente. Del mismo modo que los modelos matemáticos, los algoritmos cognitivos los tenemos que *inferir* a partir de un análisis minucioso de las interacciones entre el organismo y su medio.

Parece que la distinción se tiene que establecer en un nivel puramente conceptual. Para llevar a buen puerto esta tarea, hay un criterio que nos puede resultar de suma utilidad: en virtud de aquello que representan, los modelos psicológicos son, por definición, más generales que los algoritmos cognitivos. En este contexto, *modelar un algoritmo* sería equivalente a *investigar un algoritmo*. El propio Marr (1983/2010) ya es consciente de que

con los algoritmos pasa algo que Cartwright (1980/1991) había descripto previamente para las leyes y que yo extendí a toda formalización científica: éstos pueden ser redundantes. Es decir, diferentes postulaciones del un algoritmo puede dar cuenta igual de bien de los datos empíricos disponibles. De este modo:

Elegir entre un determinado algoritmo y otro depende de algunas características particularmente deseables o indeseables que el algoritmo pueda tener; por ejemplo, un algoritmo puede ser más eficiente que otro, u otro puede ser ligeramente menos eficiente pero más robusto [...]. O, una vez más, un algoritmo puede ser paralelo y otro serial. La elección depende, entonces, del tipo de hardware o mecanismo en que el algoritmo está incorporado físicamente (Marr, 1983/2010, pp. 23-24).

Como ha de ser claro, y a diferencia de las leyes y de los modelos científicos, la elección de un algoritmo (aunque está sea provisional) sí resulta obligatoria, dado que ésta es un aspecto central de la reconstrucción causal de los fenómenos cognitivos.

Ahora, habría de ser claro que los algoritmos entran en juego en momentos puntuales del procesamiento de información y puede ser vinculado con elementos específicos de la arquitectura del sistema. Los modelos psicológicos formales, por su parte, no pretenden dar cuenta de transiciones entre representaciones, sino de *procesos* cognitivos mucho más amplios. Un modelo psicológico podría incluir entre sus partes constitutivas uno o más algoritmos cognitivos (dependiendo de cuan específico sea) pero

también —y de vital importancia— algunos de sus parámetros deben dar cuenta de la estructura del ambiente en que actúa el organismo (*i.e.*, no referir a computaciones).

Puede que uno de los errores más graves y censurables de la ciencia psicológica en su conjunto tenga que ver con su negligencia (bastante generalizada) a la hora de investigar la manera en que nuestras (limitadas) capacidades cognitivas interactúan con el mundo q más allá de nuestros cerebros para permitirnos emitir conductas inteligentes (Clark, 1997). En sus intentos por igualar algoritmos con modelos psicológicos, el AEC está cayendo en la tentación de hacer de los procesos cognitivos procesos computacionales totalmente internos y privados. Sostengo que una ciencia de la mente que esté basada el cálculo de probabilidades, el ordenamiento de preferencias y una toma óptima de decisiones está irremediabilmente condenada al fracaso. El delineamiento de una alternativa naturalista viable a esta posición *internalista-formalista-intelectualista* es el objetivo de la segunda parte de este trabajo.

A lo largo de este capítulo he intentado perfilar a grandes rasgos los que, considero, son los pilares metodológicos y conceptuales del AEC. Uno de los referentes más directos de esta tradición se encuentra en la adopción de los modelos matemáticos en economía desde el siglo XIX. He argumentado, siguiendo a Morgan (2009) que esta transición de las descripciones verbales de los fenómenos al uso de modelos formales conlleva una transformación radical en el *estilo de razonamiento* del que echan mano los

científicos y que, contrario a aquellos que defiende que el objetivo último de la ciencia es el desarrollo de expresiones matemáticas, éstas sólo tienen sentido como parte de un entramado de prácticas explicativas mucho más amplio y complejo (Cartwright, 1980/1991). En este contexto, la pugna sostenida por los defensores de la teoría computacional clásica y los analistas experimentales de la conducta con respecto a qué característica debe poseer una descripción naturalista de los fenómenos mentales carece completamente de sentido: la cuestión no está en si las ciencias cognitivas deben o no hacer uso de descripciones verbales o de expresiones lógicas y matemáticas para formular sus teorías, sino en determinar *dónde* son más útiles cada una de estas herramientas epistemológicas.

SEGUNDA PARTE:

RAZÓN Y LÓGICA: HACIA UNA TEORÍA (REALMENTE) NATURALISTA

DEL RAZONAMIENTO HUMANO

CAPITULO CINCO

PRELIMINARES PARA UNA TEORÍA NATURALISTA DEL CONOCIMIENTO HUMANO

We use intelligence to structure our environment so that we can succeed with less intelligence. Our brains make the world smart so that we can be dumb in peace!

Andy Clark.

El trasfondo teórico que he reconstruido a la largo de esta tesis es fundamental para comprender la forma en que los fenómenos psicológicos son entendidos en la ciencia cognitiva contemporánea. En los dos capítulos anteriores presenté un marco conceptual donde la noción de cognición es simplemente indisociable de la de computación. La idea básica es que se puede especificar una *arquitectura simbólica* que dé cuenta del funcionamiento mental a partir de las reglas que guían el procesamiento de la información —*i.e.*, la transición de un tipo de representaciones a otro— (Pylyshyn, 1989/2007).

Con el trabajo de Marr (1983/2010), la perspectiva simbólico-representacional se muy fortalecida al incorporar de lleno el nivel físico (*i.e.*, neurofisiológico) como un elemento central de este esquema explicativo. De este modo, la mente-cerebro es entendida como una computadora y el objetivo

último de la ciencia psicológica pasa a ser la reconstrucción de código en la que está programada —más adecuado sería decir que, con la popularidad de la tesis de la modularidad masiva (Sperber, 1994), la mente-cerebro pasa a ser vista como *un conjunto* de computadoras relativamente independientes (*i.e.*, con arquitecturas simbólicas diferentes), “programadas” a lo largo de nuestro desarrollo como especie para hacer frente a problemas con una alta significancia adaptativa (Buss, 2008; Cosmides y Tooby, 1994; 1997; Pinker, 1997).

No obstante, en sí mismas las aproximaciones computacionales permanecen mayormente silentes en lo que tiene que ver con dos problemas fundamentales de la ciencia cognitiva: (1) el *problema del conocimiento* (Pozo, 2003) y (2) el problema de la *racionalidad* (Bermúdez, 2003; Hurley y Nudds, 2006, y los trabajos ahí compilados). Incluso tomando al pie de la letra los criterios propuestos por Turing (1950) para la adscripción de conducta inteligente, una teoría sobre la naturaleza computacional de los procesos neurofisiológicos y cognitivos no es en sí misma una teoría sobre la racionalidad ni, mucho menos, una teoría sobre el conocimiento (véase el capítulo 1 en Hurley y Nudds, 2006 para una discusión de las diferencias entre inteligencia y racionalidad).

Sin embargo, esta visión de la cognición como computación resulta un impuso muy grande para la legitimización de una concepción de la racionalidad profundamente arraigada desde (cuando menos) el *Discours de la méthode* de Descartes (1637/2001): la idea de que ésta es, en esencia, una cuestión de *pensar a partir de las reglas de la lógica y matemática*. La

cuestión aquí ya no es sólo que los procesos cognitivos puedan ser especificables a partir de la lógica y la matemática, sino que se defiende la idea de que pensar “correctamente” consiste en la *aplicación* de principios lógicos y matemáticos.

Descartes (1637/2001) establece una división tajante entre el nivel teórico y el nivel práctico de la experiencia, considerando que en nuestra interacción diaria con las cosas y las personas nos vemos muchas veces obligados a guiar nuestro comportamiento por elementos extrateóricos, como el prejuicio y la tradición. En el nivel intelectual, no obstante, un *ejercicio sistemático de la razón* libre de prejuicios nos permitiría acceder a las verdades *claras y distintas* que pueden servirnos como fundamento sobre el cual edificar el edificio entero de nuestro conocimiento. La razón omnisciente cartesiana es en muchos sentidos la marca misma de la modernidad. Otto Neurath (1913/1983) ya criticaba el “pseudoracionalismo” imperante entre sus contemporáneos (incluidos los otros positivistas lógicos) al considerar que, a diferencia de Descartes, habían extendido la exigencia de que nuestras acciones estuvieran gobernadas por la razón teórica a todos los niveles de nuestra experiencia... ¡en muchos sentidos eran más racionalistas que el mismísimo Descartes!

Neurath (1913/1983) rechaza por igual la distinción entre nivel práctico y nivel teórico de la experiencia y el supuesto de que es posible un razonamiento puro, libre de todo condicionamiento. Para él, toda nuestra actividad mental está mediada por nuestras ideas sobre el mundo: “No hay *tabula rasa*. Somos como navegantes que tienen que reparar su nave en

pleno mar abierto, incapaces de atrancarla en un dique seco para reconstruirla con los mejores materiales” (Neurath, 1932/1983, p. 92). Sólo hoy, a casi cien años de haber sido originalmente formuladas, las ideas Neurath (1913/1983; 1932/1983) han comenzado a recibir la atención que merecen.

Al finalizar el capítulo 3 argumenté que de la teoría simbólico-representacional se desprende una visión de la cognición que es lingüística en dos sentidos profundamente interrelacionados, pero no coextensivos: entendidos como fenómenos computacionales, los procesos cognitivos pueden ser descritos en términos lingüísticos, pero también (y de forma muy importante) tanto la noción de *agente racional* como la de *conocimiento* pueden ser entendidas a partir de los mismos conceptos empleados para hablar de los lenguajes formales. En esta segunda sección del trabajo me concentraré en el segundo elemento de esta distinción. Para ello, introduzco una descripción de los procesos cognitivos humanos que, sin equiparar la naturaleza del lenguaje con la del pensamiento y el conocimiento —*i.e.*, trascendiendo el supuesto de que los tres son fenómenos sintácticamente especificables—, nos permita dar cuenta de la manera en que un sistema tan sofisticado y complejo como lo es el conocimiento científico puede estar soportado en las capacidades cognitivas de los miembros de nuestra especie, mucho menos regidas por los principios de la lógica y la matemática de lo que tradicionalmente se ha venido aceptado.

5.1. RAZÓN Y LÓGICA

Recordemos que Brun y Kuenzle (2008) sostienen que, en el desarrollo de las discusiones sobre la forma en que nuestro conocimiento están soportados por nuestras competencias psicológicas: “las evaluaciones epistémicas de proposiciones, sentencias o estados mentales fueron realizadas primero y entonces los agentes, actos y procesos epistémicos fueron explicados en términos de estos análisis” (p. 6). De este modo, la caracterización de la ciencia en términos de un sistema de proposiciones lógicamente relacionadas sobre el mundo es previa a cualquier descripción de los procesos (individuales y colectivos) que permiten su construcción y, aún más, a la delimitación de las capacidades cognitivas de los agentes los llevan a cabo.

Como hacen ver Brun y Kuenzle (2008), una de las consecuencias de que la especificación (por demás arbitraria) del producto (el conocimiento científico) haya antecedido a la descripción del proceso y de los agentes que posibilitan su construcción y desarrollo es que éstos últimos terminaron por asimilar como propias las características que se consideran predicables del primero. Siendo la lógica la medida del conocimiento científico, resulta casi natural pensar que también habría de serlo de los agentes y procesos epistémicos.

La “revolución cognitiva” no es más que un intento de dar cuenta de los fenómenos psicológicos en términos que se ajusten —y en última instancia posibiliten— los procesos de deducción lógica que los epistemólogos positivistas consideraban indispensables a la hora de dar cuenta de la justifi-

cación del conocimiento. Con la hipótesis del “lenguaje del pensamiento”, Fodor (1975) completa esta imagen al sostener que, en virtud de la naturaleza computacional de la cognición, el proceso global de pensamiento (tanto humano como animal) sólo puede ser entendido a partir de la postulación de un código interno y privado por medio del cual los organismos *se representan* los objetos y las relaciones del mundo. Así pues, la hipótesis del “lenguaje del pensamiento” ya no es sólo una propuesta sobre el funcionamiento mental, sino que —dando por hecho su naturaleza simbólica— eleva la manipulación sistemática de símbolos como el criterio último de cualquier adscripción de racionalidad.

No obstante, y como argumenté en el capítulo 2, el supuesto según el cual el dominio de un lenguaje (sea innato o sea adquirido) es un prerrequisito para cualquier proceso de pensamiento trasciende ampliamente el contexto de la teoría computacional de la mente. Ideas similares ya habían sido defendidas por autores tan promitentes como Lev Vygotsky (1934/1986) o Jean Piaget (1928/2002) desde una tradición intelectual que, supuestamente, representa la antítesis de las teorías formalistas del funcionamiento mental. Lo que es más, la intuición “lenguaje = pensamiento” es uno de los pilares sobre los que descansa todo nuestro sentido común. En términos prácticos, lo único que Fodor (1975) hace es extender al nivel de la racionalidad animal una caracterización que ya era ampliamente aceptada en el contexto del pensamiento humano.

Estos supuestos contribuyeron a consolidar la idea de que *conocer* es esencialmente una actividad individual y privada que tiene lugar a partir de

la aplicación de unas cuantas reglas infalibles y universales de razonamiento a una serie de proposiciones sobre el mundo en virtud de su forma lógica y con independencia del contexto de su formulación. El lenguaje es visto como el vehículo por excelencia para el pensamiento (Fodor, 1975), con lo que, dada esta equiparación entre *racionalidad práctica* y *racionalidad teórica*, para dar cuenta de las *causas* de la conducta de un agente uno no tiene más que cuestionarlo por sus *razones* (Davidson, 1963/2001).

Entre los problemas más evidentes con esta aproximación al estudio de los fenómenos mentales se pueden mencionar su inadecuación a la hora de intentar explicar el comportamiento y la cognición de animales carentes de lenguaje y lo complicado que resulta trazar la continuidad entre el pensamiento de los organismos *alingüísticos* o *prelingüísticos* y aquel de los humanos lingüísticamente competentes (Bermúdez, 2003). Sin embargo, los intentos por formalizar el estudio de la cognición a partir de la lógica o la estadística también se han quedado bastante cortos a la hora de caracterizar la “cognición lingüística” (*i.e.* humana).

5.2. RAZÓN, MÁS ALLÁ DE LA LÓGICA Y LA MATEMÁTICA

No deja de resultar fascinante que en ciencia cognitiva el investigador que con gran esmero explora la naturaleza de los fenómenos psicológicos sea al mismo tiempo parte de aquel segmento de la naturaleza de busca comprender. Como con ningún otro dominio de la empresa científica, los resultados surgidos de la investigación del razonamiento comprometen la imagen que

tenemos de nosotros mismos y de nuestro lugar en el mundo. Como Gerd Gigerenzer y Henry Brighton (2008) lo comentan:

El Antiguo Testamento dice que Dios creó a los seres humanos a su imagen y entonces los dejó dominar sobre todos los demás animales, de los que diferimos fundamentalmente (Génesis 1:26). No resulta enteramente accidental que en ciencia cognitiva una cierta omniscencia (el conocimiento de todas las probabilidades y utilidades relevantes, por ejemplo) y cierta omnipotencia (la habilidad para computar funciones complejas en una fracción de segundo) hayan definido nuestros modelos de la cognición (p. 108).

De este modo: “El razonamiento correcto habría de adherirse a las leyes de la lógica, el cálculo de probabilidades o la maximización de la utilidad esperada; de lo contrario, debería haber algún defecto cognitivo o motivacional” (Gigerenzer, 2004; p. 62). Como he argumentado, el desarrollo de disciplinas como la economía o las ciencias cognitivas es simplemente incomprensible sin apelar de forma directa a esta noción de *racionalidad*.

En primera instancia, la evidencia empírica acumulada durante décadas de investigación en ciencia cognitiva terminó por obligarnos a aceptar que existe una gran continuidad entre la cognición animal y la humana por lo que, o aceptábamos la atribución a otros animales de procesos de razonamiento análogos a los que predicamos de nosotros mismos (la estrategia asumida por Fodor, 1983), o renunciábamos a la idea de que nuestra conducta está sustentada en procesos plenamente racionales. A riesgo de sonar categórico, bien podría decirse que, en las últimas cuatro décadas, las cien-

cias cognitivas han estado atrapadas en esta disyunción, aunque con cierta tendencia hacia la defensa del estatus racional del comportamiento.

Si bien es cierto que desde hace años se acepta que la nuestra cognición muchas veces se aleja de los criterios de racionalidad impuestos por la lógica y la estadística (véase, por ejemplo, Wason y Shapiro, 1971), éstas no han dejado de jugar un papel fundamental en la conceptualización y evaluación de los procesos cognitivos.

Una vez asimilado este shock intelectual (razonamiento \neq lógica), el estudio del pensamiento pasó a desenvolverse dentro de supuestos defendidos por lo que se ha dado en llamar el “programa de heurísticos y sesgos”, desarrollado por Amos Tversky y Daniel Kahneman (1974, Kahneman y Tversky, 1984). Desde esta perspectiva, las “desviaciones” de nuestro razonamiento respecto de la lógica y la probabilidad representan “errores” que algunas veces pueden llevarnos a tomar decisiones subóptimas y poco adaptativas pero que, “por fortuna”, serían susceptibles de ser superados a través de la instrucción adecuada (Sutherland, 1996). Así entendido, el problema central de la investigación del razonamiento y el comportamiento racional no es uno de adecuación, sino de aplicación: si el comportamiento de los organismos se alejan de lo predicho por la lógica y la estadística, ¡peor para los organismos!

No obstante, en la última década Gerd Gigerenzer y sus colaboradores (2004; Gigerenzer y Brighton, 2008; Gigerenzer y Gaissmaier, 2011; Todd y Gigerenzer, 2000) han abordado el estudio de los heurísticos y sesgos desde una perspectiva muy diferente: lo que se ha dado en llamar “racio-

nalidad limitada” (*bounded rationality*) o “racionalidad ecológica”. Todd y Gigerenzer (2000) argumentan que, *cuando se toma en cuenta la estructura del ambiente*, reglas simples de decisión como los heurísticos superan muchas veces el desempeño predicho por complejos modelos estadísticos.

De este modo, no hay nada de malo con nuestra forma cotidiana de razonar ni razones para suponer que nuestros procesos cognitivos deberían de ajustarse a los criterios prescritos por sistemas históricamente desarrollados y culturalmente específicos como son la lógica o la estadística (Weinberg, Nichols y Stich, 2001/2008; Machery, Mallon, Nichols y Stich, 2004/2008). Esto es lo que uno debería esperar si se acepta que la cognición es un producto más de la evolución biológica, donde la capacidad de dar respuestas flexibles de forma rápida y frugal debió haber sido de gran importancia adaptativa (Gigerenzer y Brighton, 2008).

Detrás de esta distinción entre “procesos plenamente racionales” y heurísticos y sesgos de razonamiento se haya la cuestión de fondo con respecto a cuál es la mejor forma de caracterizar los procesos *biológicos* de toma de decisiones. Siguiendo las consideraciones de Bermúdez (2003), pensemos que estamos investigando la conducta de un organismo (O) ante la importante cuestión de conseguir alimento; para ello, O distribuye su tiempo de búsqueda entre diferentes sitios de aprovisionamiento de forma tal que el patrón global de conducta queda capturado por dos sistemas explicativos diametralmente distintos: (1) el primero asume la existencia de procesos de *racionalidad máxima*, de tal forma que, cada vez que llega a un nuevo sitio de aprovisionamiento ($S_1, S_2, S_3 \dots S_n$), O realiza una eva-

luación exhaustiva de *todos* los aspectos del medio ambiente que resultan relevantes para determinar la disponibilidad de alimento; si de lo anterior resulta que existe una alta probabilidad de encontrar una buena cantidad de nutrientes en S_1 , O determina el tiempo (T) que dispondrá ahí antes de moverse a un S_2 , donde este proceso comenzaría de nuevo; (2) la explicación en términos de *racionalidad limitada* supone que O cuenta con dos reglas simples que guían su conducta de forrajeo: a) Si después de 1 minutos de haber llegado a un sitio de aprovisionamiento, S_1 , no se ha obtenido alimento alguno, muévete a un S_2 ; y b) muévete a un S_2 si el tiempo transcurrido desde la última porción de alimento recuperado supera los 20 segundos. Es importante destacar que en éste último sistema explicativo a) y b) cumplen plenamente con los criterios de rapidez y frugalidad que Gigerenzer considera defnitorios de un proceso heurístico de toma de decisiones: (i) la implementación de ambas reglas es hasta tal punto rápida que su selección y aplicación coinciden temporalmente —es decir, el mismo flujo de la conducta presente condiciona a la conducta futura, sin necesidad de apelar proceso alguno de deliberación previa; y (ii) ambas reglas son frugales dado que el tiempo sin obtener comida resulta la única variable importante para determinar el flujo comportamental, quedando de lado otras fuentes de información relevante para la disponibilidad de alimento en un sitio de aprovisionamiento, como podría ser la calidad del suelo, el tipo de vegetación o la presencia de otros animales (por mencionar sólo algunas).

Suponiendo que el poder predictivo de (1) y (2) sea muy similar y que no dispongamos algún dato empírico concluyente para inclinarnos por

una explicación o la otra, la pregunta sería, ¿cuál de las dos habríamos de elegir? Varias consideraciones entran en juego en este punto. Primero, pensemos que nos damos a la tarea de formalizar ambas explicaciones de tal manera que cada una quede expresada por una ecuación, Ec_1 y Ec_2 , respectivamente; habría de resultar claro que, en tanto modelos, Ec_1 y Ec_2 son *redundantes* en el sentido expuesto por Cartwright (1980/1991). Así pues, un buen modelo matemático sobre la cognición no es sólo aquel que da cuenta de los datos, sino también uno que nos cuenta una *buena historia* de esos datos.

Ahora bien, ¿qué historia resulta más convincente? Un criterio de cual se suele echar mano en los estudios de psicología comparada es el formulado por C. Lloyd Morgan (1894, citado en Shettleworth, 2010, p. 17) en su famoso canon: “en ningún caso podemos interpretar una acción como el resultado de la existencia de una facultad psicológica superior si puede ser interpretada como resultado del ejercicio de una que está en un nivel más bajo en la escala psicológica”. Sin embargo, como Sara Shettleworth (2010) hace ver, no siempre queda claro cómo interpretar “superior” e “inferior” en la “escala psicológica”. ¿Hay alguna forma inteligible en la que podamos comprender esta exigencia?

5.2.1. LO BÁSICO Y NO TAN BÁSICO DEL APRENDIZAJE ASOCIATIVO

Existe una fuerte tendencia a considerar como “básico” todo aquel comportamiento que pueda ser explicado a partir de mecanismos de aprendizaje

asociativo y/o sesgos perceptuales o conductuales típicos de una especie (Shettleworth, 2010) y como “superior” todo comportamiento que no sea explicable en estos términos (de forma típica, aquellos cuya explicación requiere la apelación a procesos inferenciales). No obstante, esta propuesta parece subestimar la complejidad de la organización de la conducta; como Penn y Povinelli (2007) comentan:

La venerable dicotomía entre asociacionismo y explicación inferencial de la cognición causal no-humana es tanto engañosa como improductiva. [...] La situación real parece ser mucho más complicada, multifactorial y fascinante. [...] Muchos aspectos del aprendizaje causal humano y no-humano son, por igual, parsimoniosamente explicados en términos de alguna forma de condicionamiento asociativo. Por otra parte, tanto animales humanos como no-humanos son sensibles a restricciones específicas de relaciones causales *sensu strictu* [...] Esto implica que la inducción causal no es simplemente reducible al aprendizaje de contingencias, tanto en sujetos humanos como en los no-humanos. Incluso los animales no-humanos emplean mecanismos cognitivos que distinguen entre causalidad y covariación (p. 110).

Esta consideración resulta fácilmente extrapolable a cualquier otro dominio de la investigación de la cognición biológica (cuando menos en mamíferos), sobre todo si consideramos la prevalencia de los procesos de evaluación de contingencias en todos los niveles de la organización cognitiva. Dicho de otro modo, para establecer una división tajante entre sistemas “básicos” de aprendizaje asociativo y mecanismos de cognición “superior”, primero es necesario demostrar que existen amplias parcelas de la vida mental de los

organismos cuya explicación puede ser agotada *exclusivamente* a partir de las herramientas teóricas de la psicología asociacionista clásica.

Considero que el estado actual de la investigación dentro del análisis experimental de la conducta nos obliga a avanzar otra alternativa, ya no más radical, sin sólo más empíricamente plausible: una vez que los mecanismos neurofisiológicos que soportan las diferentes formas de condicionamiento se consolidaron como elementos constitutivos del tejido neural —evento que debió tener lugar muy temprano en el desarrollo del sistema nervioso, como lo demuestra el hecho de que los mayores progresos en el estudio de los sustratos neurales del aprendizaje asociativo vengan del estudio de organismos invertebrados, tanto en el caso del condicionamiento clásico (Kandel, 2008), como en el del condicionamiento operante (Brembs, 2003)—, éstos se constituyeron como los cimientos sobre los que se erigieron el resto de los mecanismos cognitivos. Así pues, no se podría hablar de que existan conductas “básicas” en el sentido en que éstas sean consecuencia exclusiva de algún proceso de condicionamiento, dado que no sería posible hablar de *aprendizajes asociativos puros* —entendidos éstos en su forma más tradicional como “una asociación que simplemente conecta la imagen mental de un par de eventos en tal forma que la activación de una imagen causa la activación (o inhibición) de la otra” (Shanks, 2007, citado en Mitchell, De Houwer y Lovibond, 2008).

La idea de que los mecanismos involucrados en el aprendizaje asociativo son simplemente indisociables de otros sistemas cognitivos se ve apoyada por la abundancia de resultados empíricos que indican que, más que

el sólo establecimiento de una asociación simple y automática *entre* estímulos, el estudio de los procesos de condicionamiento —particularmente, de condicionamiento instrumental— revela que los organismos son mucho más sensibles a la estructura del ambiente de lo que las teorías clásicas del aprendizaje hubieran estado dispuestas a aceptar. De este modo, Rescorla y Skucy (1969, citado en Bermúdez, 2003) mostraron de forma convincente que las ratas entrenadas para presionar una palanca a cambio de comida dejaron de emitir la conducta instrumental cuando ésta dejó de estar *causalmente* vinculada con la ocurrencia del reforzador (que se liberaba con total independencia del comportamiento del animal).

Haciendo uso de diseños experimentales aún más sofisticados, los analistas de la conducta han descrito una amplia gama de fenómenos que ponen en entredicho la existencia de un sistema “básico” para la formación de links que permitan la coactivación automática (excitatoria o inhibitoria) de las representaciones de estímulos (véase Penn y Povinelli, 2007 para una revisión de la evidencia empírica y argumentos interesantes en este sentido). Un ejemplo simple es el llamado bloqueo hacia atrás (*backward blocking*), donde se entrena a ratas para responder ante un estímulo compuesto (AX+), posteriormente se presenta A+, lo cual bloquea la respuesta al estímulo X en posteriores ensayos de prueba. Tal parece que los animales “infirieran” que, dado que A+ da cuenta, por sí mismo, de la ocurrencia del reforzador, el estímulo X realmente no estaría señalando ningún acontecimiento relevante (Penn y Povinelli, 2007; Shettleworth, 2010). Otro ejemplo ilustrativo es la demostración de Denniston, et al. (2003, citado

en Penn y Povinelli, 2007) de la existencia de “inferencias transitivas” en animales no-humanos: se entrena a la rata ante un estímulo compuesto AX+ y posteriormente ante otro XY+; los animales que fueron sometidas a ensayos A- de extinción disminuyeron su tasa de respuesta ante el estímulo Y, incluso aunque A y Y nunca ocurrieron juntos.

Es importante aclarar que, con la suficiente paciencia e imaginación, cualquier conducta puede ser explicada en términos “puramente asociativos” o “puramente inferenciales” —este último caso queda perfectamente reflejado en el “modelo teórico de decisión del razonamiento práctico” propuesto por Fodor (1975). En el caso del análisis experimental de la conducta, existe una tendencia a modificar los supuestos de los que parten sus modelos del aprendizaje a medida que surgen más y más datos que muestran su insuficiencia. Al respecto Mitchell, De Houwer y Lovibond (2008), comentan:

Al tiempo que modelos individuales como el de Rescorla-Wagner son muy parsimoniosos, la clase entera de teorías que supuestamente describen la forma en que los links son formados, no lo es. Aunque extender los modelos en una manera post hoc no es, en principio, problemático, evaluar el modelo extendido sólo de cara a un conjunto de datos (para el que la extensión fue específicamente diseñada) es peligroso. La generalización del nuevo modelo a otros conjuntos de datos debería ser demostrada; de lo contrario, se corre el riesgo de que un diferente modelo basado en la formación de links entre estímulos sea generado post hoc para dar cuenta de cada resultado experimental observado (p. 195).

Es importante destacar que Mitchell, De Houwer y Lovibond (2008) están particularmente interesados en argumentar a favor de “la naturaleza proposicional del aprendizaje asociativo humano” (que es, precisamente, el título de su trabajo); no obstante, sus comentarios dejan claro que sus pretensiones son mucho más amplias:

No queremos argumentar que los humanos poseen un sistema cognitivo de aprendizaje único. Esta postura implica que el aprendizaje de los animales no-humanos es también un proceso de adquisición de creencias. Por lo tanto, el complejo sistema representacional del que disponemos como especie evolucionó de los sistemas cognitivos similares, pero más simples, de nuestros ancestros; y muchas diferencias observadas entre el aprendizaje humano y el no-humano son cuantitativas, no cualitativas (p. 196).

Como Penn, Holyoak y Povinelli (2008) hacen ver, esta venerable tradición intelectual se remonta al mismísimo Darwin y ha jugado un rol muy importante en el desarrollo de nuestras concepciones sobre la cognición y el comportamiento. En términos generales, concuerdo con Mitchell, De Houwer y Lovibond (2008) cuando argumentan que los recursos conceptuales de las teorías clásicas del aprendizaje resultan por demás insuficientes para dar cuenta de los datos surgidos en esta área de la investigación psicológica. Pero no puedo estar más en desacuerdo cuando pretenden que de este hecho se sigue que la única forma de comprender el aprendizaje o cualquier otro fenómeno cognitivo sea en términos proposicionales.

Tal parece que Mitchell, De Houwer y Lovibond (2008) están cayendo en la misma tentación que cayera Fodor (1975) hace más de 30 años: evidenciar lo limitado que resulta el asociacionismo para explicar cualquier nivel de la cognición *no equivale a demostrar la conveniencia de una posición que entienda que todo proceso cognitivo es de naturaleza inferencial y depende de información representada proposicionalmente*. Este *falso dilema* resulta por demás falaz: entre uno y otro extremo hay una gran gama aproximaciones teóricas que merecen ser exploradas; y si bien es cierto que aún estamos lejos de una solución concluyente en este sentido, la evidencia empírica disponible es más que suficiente para rechazar las posturas asociacionistas e inferenciales extremas.

Resumiendo, resulta muy difícil sostener una distinción entre procesos asociativos “básicos” y mecanismos de cognición “superior” dado que los fenómenos psicológicos suelen ser el resultado de funcionamiento coordinado de diferentes sistemas cognitivos. No obstante, es igualmente injustificado defender que la cognición es un hecho unitario dependiente de algún tipo de sistema general de procesamiento, sea que este se piense en términos puramente asociativos o puramente inferenciales. Por el momento dejaré abierta la cuestión de cuál es (al menos desde mi punto de vista) la mejor forma de definir la relación que establecen los diferentes sistemas cognitivos para revisar otra forma en que podría entenderse la exigencia impuesta por el canon de Morgan.

5.2.2. SISTEMAS DUALES DE RAZONAMIENTO: EVOLUCIÓN Y COGNICIÓN

Muy relacionada con la posibilidad anterior, pero no equivalente, es la idea de que la “escala psicológica” puede ser entendida en términos evolutivos, donde “básicos” serían los sistemas cognitivos que surgieron primero en escalar filogenético, mientras que “superiores” serían aquellos que entraron en escena de forma más reciente. Ha de ser claro que la misma crítica hecha a la posición anterior es aplicable en este caso: no hay nada en el hecho de que un determinado sistema sea evolutivamente antiguo que nos lleve a afirmar (o siquiera suponer) que esté dissociado del resto de los mecanismos cognitivos.

Con todo, en la última década han comenzado a proliferar las explicaciones del razonamiento en términos de “sistemas duales” (Evans, 2003; 2008). Como el propio término lo indica, la idea básica detrás de estas teorías es que existen dos sistemas básicos y relativamente independientes de razonamiento, cada uno con su propia historia evolutiva. En palabras de Jonathan St. B.T. Evans (2003):

Los investigadores del pensamiento y el razonamiento han propuesto recientemente que hay dos sistemas cognitivos distintos que subyacen al razonamiento. El sistema 1, antiguo en términos evolutivos y que es compartido con otros animales: comprende un conjunto de subsistemas autónomos que incluyen tanto módulos de *input* innato como conocimiento específico de dominio adquirido por medio de mecanismos de aprendizaje de dominio general. El sistema 2 es evolutivamente reciente y distintivamente humano: permite el razonamiento abstracto

y el pensamiento hipotético, pero está constreñido por la memoria de trabajo y correlacionado con medidas de inteligencia general (p. 454).

La posición criticada por Mitchell, De Houwer y Lovibond (2008) ya es plenamente una teoría dual del razonamiento, dado que propone la existencia de *un* sistema “básico” para el aprendizaje asociativo —de naturaleza “inconsciente, rápida, automática y poseedor una amplia capacidad”— y *otro* que subyacería a los procesos cognitivos “superiores”—dependiente del lenguaje y cuya puesta en marcha sería “consciente, lenta y deliberada” (Evans, 2008, p. 256).

Esta descripción nos revela uno de los mayores problemas de las teorías duales del razonamiento: si nos tomamos la propuesta en serio e intentamos ser lo suficientemente específicos como para caracterizar uno de los dos sistemas de forma tal que resulte tanto conceptual como empíricamente abordable, de manera inmediata el otro se vuelve demasiado general e impreciso. De este modo, los mismos criterios que nos llevaron a plantear una distinción entre dos diferentes sistemas de razonamiento nos comprometerían a querer aclarar la esquivada naturaleza del sistema cuya descripción hayamos dejando para un segundo momento. Esto nos facultaría para postular que a su interior existe una serie de “subsistemas”, algunos de los que guardaría una gran similitud funcional con aquel sistema del que supuestamente son contraparte, ¡diluyéndose así la supuesta distinción entre dos grandes sistemas de razonamiento! (véase Evans, 2008, para una revisión de las discusiones contemporáneas en esta área). Con esto, las teorías

duales del razonamiento no hacen más que confirmar lo que ya sabíamos: que la cognición es el resultado de *funcionamiento coordinado de una serie de sistemas* y que el gran reto está, precisamente, en determinar cómo se *relacionan* unos con otros para permitirnos emitir comportamiento inteligente y adaptarnos a nuestro entorno.

Plantear dicotomías entre un sistema “básico” para el aprendizaje asociativo y otro “avanzado” para la cognición “superior” (que vendría siendo todo lo que no sea aprendizaje asociativo) o entre un sistema 2, lógico y deliberado, y un sistema 1, inconsciente y automático (que en última instancia sería todo aquello más allá de una inferencia verbalmente expresable), resulta un total sinsentido.

El canon de Morgan sólo es comprensible si la explicación de un determinado fenómeno psicológico puede ser *agotada* a partir del funcionamiento de un determinado mecanismo cognitivo, claramente discernible del resto de la arquitectura mental, y en la medida en que los diferentes sistemas cognitivos pueden ser ordenados a partir de un criterio objetivo (en este caso, su origen evolutivo). Considero que no hay razón alguna para pensar que alguna de estas dos condiciones pueda ser cabalmente cumplida.

De este modo, el canon de Morgan sólo puede ser salvado si se considera como una apelación general al criterio de parsimonia *aplicable a nuestras explicaciones sobre la cognición y no como un principio sobre la naturaleza de los procesos que las motivan*. Así pues, si contamos con dos explicaciones que den cuenta *por igual* de la evidencia empírica disponible —hecho que de entrada considero poco probable, aunque no me voy a

detener a discutir esta afirmación aquí— habríamos de elegir aquella que eche mano de menos supuestos y que postule procesos más simples.

Podríamos decir que en ciencia las explicaciones complejas son “inocentes” (no capturan las *causas* detrás de un determinado hecho) “hasta que se demuestre lo contrario” (aunque esto no equivale a defender la adecuación de una determinada descripción por el sólo hecho de ser simple). De este modo, *si hemos de postular “reglas de razonamiento”* (y me gustaría enfatizar el carácter condicional de este supuesto), habíamos de preferir aquellas que sean más simples y empíricamente contrastables... bien parece que el razonamiento de O podría muy bien ser de naturaleza heurística, ¡al menos que O quiera *argumentar* lo contrario!

5.3. LA NATURALEZA INTERACTIVA DE LOS SISTEMAS COGNITIVOS

Pero entonces, ¿cuán es la forma más productiva de comprender y explorar la forma en que los diferentes sistemas cognitivos interactúan para dar lugar a nuestra vida mental? Considero que aún es necesario mucho trabajo y discusión para aclarar esta importante cuestión. No obstante, pienso que un buen punto de partida es la estrategia asumida por Antonio Damasio (1994/2006) en su intento por dilucidar las relaciones existentes entre el procesamiento emocional y el razonamiento. La propuesta, en sí misma, ya es de suma relevancia para la cuestión que aquí vengo discutiendo, dado que muestra de manera particularmente convincente cuán incongruente es seguir sosteniendo la tradicional dicotomía que contrapone razón contra

emoción: las personas cuya percepción emocional está alterada presentan también profundas perturbaciones en sus procesos de toma de decisiones (sobre todo en el terreno personal) y en su interacción social, lo cual los lleva a tomar rutas de acción particularmente autodestructivas, ¡aún cuando otros aspectos de su vida intelectual se mantengan intactos y obtengan coeficientes de inteligencia general incluso superiores a la media! No obstante, más que una revisión exhaustiva de la teoría del marcador somático desarrollada por Damasio (1994/2006), me gustaría enfatizar la forma en que construye su esquema explicativo.

Damasio (1994/2006) considera que la corteza prefrontal ventromedial (la parte más anterior del cerebro, que se ubica justo por encima de las órbitas oculares) juega un papel muy importante en la integración de la información emocional dentro de nuestro proceso global de razonamiento. Veamos; dos de los aspectos más destacables de las personas que han sufrido una lesión en esta parte del cerebro —muchas veces las únicas señales claras de que “algo no está bien” con ellos, de entre los pocos casos documentados de forma sistemática, claro está— es, como ya se comentó, la gran dificultad que muestran a la hora de tener que tomar una decisión y una marcada incapacidad para ajustar su conducta a las normas y convenciones sociales (Damasio, 1994/2006; Goldberg, 2002). Una interpretación intuitiva que se podría dar a este déficit es que los pacientes que lo presentan han perdido la capacidad de *representarse mentalmente* la situación cuestión, por lo que no podrían evaluarla “de forma objetiva”.

Sin embargo, las observaciones realizadas por Damasio (1994/2006) echan por la borda esta posibilidad. Él describe con gran detalle el caso de Elliot, un paciente que fuera sometido a una ablación bilateral radical de su corteza prefrontal ventromedial tras desarrollar un tumor. Elliot siempre “había mantenido un estatus social, profesional y personal envidiable” (Damasio, 1994/2006, p. 58); no obstante, después de su intervención, se mostró incapaz de organizar su conducta con miras a la consecución de objetivos globales (particularmente, tendía a perseverar en aspectos muy específicos de la tarea y era en absoluto negligente con otros de igual o mayor importancia), lo cual lo imposibilitó a continuar su vida laboral. Si antes se había mostrado cauto e inteligente en sus resoluciones personales, comenzó a tomar decisiones temerarias que comprometían su patrimonio y a relacionarse con personas “de reputación dudosa” (p. 64), lo que lo llevó a perder casi todo cuanto tenía. Elliot parecía una persona normal en todos, menos en un aspecto: mostraba un estado emocional permanentemente sereno, indiferente, como si lo que pasaba a su alrededor no lo afectara en lo absoluto.

Para explorar hasta que punto era capaz de formarse una idea clara de un problema de carácter social, Damasio (1994/2007) confrontó a Elliot con narraciones deliberadamente ambiguas, donde él debía decir lo que podría hacer un personaje que se enfrentaba ante determinada situación (por ejemplo, un hombre que tenía que decidir cómo reaccionar tras haber roto por accidente el florero favorito de su esposa). Elliot no sólo comprendió a la perfección las implicaciones de la situación, sino que fue capaz

de ofrecer un gran número de cursos de acción perfectamente adecuados para las circunstancias, acorde con la media de los sujetos sanos. De forma análoga, obtuvo resultados normales (y, en algunos casos, incluso superiores) en su capacidad para reportar las consecuencias de acciones dadas, de imaginar medios eficaces para conseguir un objetivo en contextos de interacción social y de formular juicios morales explícitos.

De forma clara, los problemas de Elliot no podían ser atribuidos a una falta de comprensión del saber social. Pero, de algún modo, él sentenciaría su propia desgracia: “al final de esa sesión, él sonrió, aparentemente satisfecho por su rica imaginación, pero agregó: «y después de todo, ¿no siempre sabría qué hacer!»” (Damasio, 1994/2006, p. 75). Resulta adecuado aceptar que esto debería ser cierto en algún grado para cada uno de nosotros, sin embargo, el caso de Elliot alcanzaba una dimensión mucho más profunda: “el fallo de Elliot parecía consistir en que era incapaz de hacer una elección apropiada, o bien, simplemente en no poder elegir” (p. 77). Su condición médica lo llevaba a considerar con gran detalle cada una de las posibilidades que se abrían ante una determinada situación, así como sus posibles resultados y consecuencias (¡tal como lo demanda la teoría económica clásica con su *homo economicus* o la ciencia cognitiva con su “agente racional”!), sin embargo, Elliot no podía poner fin a sus deliberaciones, ¡sencillamente no sabía qué elegir!

A partir de sus observaciones en Elliot y otros pacientes, así como de lo reportado en la literatura especializada en neuropsicología, neuroquímica y psicología comparada, Damasio (1994/2006) avanza su famosa “teoría

del marcador somático”. De acuerdo con él, cuando nos enfrentamos a una situación incierta no la evaluamos con la “sangre fría”, a un nivel puramente proposicional y a partir de sus consecuencias lógicas, sino que nuestro proceso de razonamiento está guiado por una serie de reacciones emocionales que “marcan” los cursos de acción disponibles a partir de sus posibles consecuencias.

De esta forma, el sólo hecho de considerar llevar a cabo una conducta “marcada” negativamente, nos llevaría a rechazarla y buscar otra alternativa; por el contrario, aquellas acciones “marcadas” positivamente se presentarían como más *adecuadas* (y por lo general lo serían), lo cual haría altamente probable que se materialicen en comportamiento manifiesto. De acuerdo con Damasio (1994/2006), en el caso de Elliot (como en el de otros pacientes que ven alterado su procesamiento emocional), aún cuando pudiera evaluar con gran detalle tanto sus posibles actos como las consecuencias que de ellos se derivarían, la falta de un “marcador somático” que le permitiera elegir entre una posibilidad y otra bloqueaba casi en su totalidad su capacidad de tomar una decisión y, lo que es más, lo llevaba a embarcarse en conductas que la mayoría de nosotros inmediatamente rechazaríamos por considerarlas inadecuadas.

Es importante aclarar que este proceso (o al menos gran parte de él) suele tener lugar más allá de cualquier control consciente. La reacción emocional (el “marcador somático”) no entra en juego como parte de una deducción o un cálculo de probabilidades, sino que es un evento fisiológico rápido e involuntario que, precisamente, lleva a una reducción drástica de

la complejidad computacional del problema, permitiéndonos actuar inteligentemente, ¡con menos inteligencia! (obviamente, entendiendo inteligencia como un proceso de reflexión deliberada y consciente). Así pues, la propuesta de Damasio (1994/2006) es, en términos generales, consistente con la idea de racionalidad ecológica desarrollada por Gigerenzer y sus colaboradores (2004; Gigerenzer y Brighton, 2008; Gigerenzer y Gaissmaier, 2011; Todd y Gigerenzer, 2000).

Entonces, dado el rol tan importante que juega la corteza prefrontal ventromedial en nuestro proceso de razonamiento, ¿estamos pues justificados a postular la existencia de un “módulo” responsable de nuestra capacidad de juicio (particularmente implicado en el terreno moral e interpersonal)? Uno de los aspectos más atractivos de la propuesta de Damasio (1994/2006) está, precisamente, en su intento por alejarse de este esquema típico de explicación en ciencias cognitivas:

De forma concreta, bien parece que en el cerebro humano hay una serie de sistemas neurales implicados estrechamente en los procesos de pensamiento orientados a una meta, lo que nosotros llamamos razonamiento, así como en la organización de las respuestas que llamamos toma de decisiones; en ambos casos, el énfasis está puesto en el dominio personal y el de las relaciones sociales. Esta misma serie de sistemas participa también en los procesos que gobiernan la expresión y la percepción de las emociones, que se refieren en parte al procesamiento de los mensajes provenientes del cuerpo (pp. 101-102).

De este modo, la corteza prefrontal ventromedial juega un papel muy importante en este proceso, pero de ningún modo lo agota. Por el contrario, esta área opera de forma coordinada con otras regiones cerebrales, como lo son, entre otras, la amígdala, la corteza frontal dorsolateral —en ambos casos de forma bilateral—, el área somatosensorial derecha y la corteza cingulada. La elección que Damasio (1994/2006) hace del término “marcador somático” también es muy reveladora en este sentido; desde su punto de vista, el funcionamiento neural es simplemente indisociable de la actividad de la totalidad del cuerpo: un “marcador somático” refiere, en esencia, a la percepción de un estado *visceral*. Esto es muy importante dado que, como comentan Bennett y Hacker (2003), más que superar el dualismo cartesiano —como se suele proclamar algunos— la neurociencia contemporánea no ha hecho más que transitar del un dualismo sustancial (mente *vs* cuerpo) a un dualismo estructural (cuerpo *vs* cerebro).

Siendo *El error de Descartes* el título del trabajo de Damasio (1994/2006), resultaría tremendamente censurable que cayera en el mismo error que critica (por ejemplo, al suponer que el cerebro es una entidad material con propiedades en todo diferentes al resto del cuerpo, sobre el cual ejerce control absoluto). Aunque es importante mencionar que Bennett y Hacker (2003) no consideran que el trabajo de Damasio esté libre del “error mereológico” (que consiste en predicar del cerebro o sus partes atributos que sólo son predicables al organismo íntegro), muchas de sus críticas van dirigidas a trabajos previos y/o pierden la esencia de la argumentación desarrollada por Damasio (1994/2006).

Pero, ¿cómo es que un determinado curso de acción queda “marcado” positiva o negativamente? Esta consideración nos lleva a apreciar con mayor claridad el carácter interactivo de la propuesta de Damasio (1994/200). De acuerdo con él:

Nacemos dotados de mecanismos nerviosos que nos permiten responder a cierta clase de estímulos por medio de estados somáticos particulares: se trata de mecanismos que subyacen a las emociones primarias. [...] Sin embargo, la mayor parte de los marcadores somáticos de los que echamos mano para tomar decisiones probablemente se establecieron en nuestro cerebro a lo largo del proceso de educación y socialización, a través del establecimiento de una conexión entre clases particulares de estímulos y clases particulares de estados somáticos. En otros términos, descansan sobre el mecanismo de emociones secundarias (pp. 229-230).

De este modo, en el proceso de establecimiento de un “marcador somático” vinculado con determinada acción, echamos mano por igual (y de manera simultánea) de una serie de sistemas que incluyen “desde” aquellos que sustentan nuestros procesos de aprendizaje asociativo, “hasta” los que nos permiten reflexionar de forma explícita y sistemática sobre los eventos del mundo —sin conceder que existan *conductas* que dependan exclusivamente de *un* sistema cognitivo.

No obstante, aunque atractiva en muchos aspectos, la propuesta de Damasio (1994/2006) deja sin explicar una cuestión muy importante: la continuidad existente entre el razonamiento de los organismos que poseen lenguaje y aquellos que carecen de él. Como él mismo lo especifica:

Los marcadores somáticos no realizan el proceso de deliberación en nuestro lugar. Ellos ayudan a que éste se lleve a cabo, destacando ciertas opciones (sean peligrosas o sean favorables) [...] La teoría de los marcadores somáticos es, por lo tanto, compatible con la idea de que, para comportarse eficazmente en el terreno personal y social, los individuos deben representarse teóricamente y de forma adecuada tanto su propia psicología como la de los otros (p. 226).

De este modo, Damasio (1994/2006) nos presenta una propuesta más plausible del *razonamiento explícito* cotidiano y de la forma en que éste nos lleva a emitir conductas adecuadas a las exigencias de nuestro entorno. ¿Quiere decir esto que todo nuestro comportamiento —o al menos todo nuestro comportamiento social— está ineludiblemente soportado por procesos de deliberación consciente? ¿Existen procesos análogos en especies diferentes de la nuestra o su conducta está gobernada por mecanismos en esencia distintos? Pienso que la propuesta de Damasio (1994/2006) ya nos ha puesto en el camino correcto, aunque hay un trecho más que nos hace falta caminar.

5.4. LAS MÚLTIPLES CARAS DE LA RACIONALIDAD

¿Qué hace que una conducta cuente como “racional”? A pesar de su prominencia en todos los niveles del estudio de la acción humana, determinar qué cuenta y qué no como “racional” no resulta para nada una tarea sencilla (véase Hurley y Nudds, 2006 para una discusión extensa de esta cuestión).

Con fines prácticos, filósofos y científicos han tendido a considerar la adscripción de racionalidad a partir de dos criterios: una conducta puede ser catalogada como racional (1) por sus consecuencias para el organismo o (2) por el proceso que la sustenta:

La conducta de un animal puede ser descrita como racional desde una perspectiva externa, sin implicar que el proceso que la genera sea racional. Por ejemplo, una conducta puede ser descrita como racional dado que maximiza la utilidad o incrementa el *fitness* genético [...] Todo animal, si busca satisfacer sus objetivos para sobrevivir y reproducirse, debe actuar en el mundo: hacer lo apropiado dadas las circunstancias y ajustar su comportamiento si las circunstancias cambian. Podemos preguntarnos si la conducta de un animal es racional en este sentido.

Pero un animal es racional en un sentido diferente si su conducta es racionalmente seleccionada en virtud de que resulta apropiada para las circunstancias. [...] Cuando preguntamos si un animal es racional, estamos interesados no sólo en si su conducta es racional, sino también en si un proceso racional explica la conducta (Hurley y Nudds, 2006, pp. 5-6).

Debe resultar claro que existen muchas conductas racionales en el primer sentido que no necesariamente lo son en el segundo. Por ejemplo, José Luis Bermúdez (2003) presenta el caso de los tropismos —entre los que él incluye los reflejos, los mecanismos innatos de desencadenamiento (*innate releasing mechanisms*) como los mecanismos de impronta, y el condicionamiento clásico. Estrictamente hablando, estas conductas no exigen explicación en términos psicológicos; lo único que requerimos es una conexión

legaliforme entre un estímulo y una respuesta: “si se encuentra el estímulo relevante, la misma respuesta emergerá” (Bermúdez, 2003; p. 116).

Bermúdez (2003) engloba estos comportamientos bajo la etiqueta de “nivel 0” de racionalidad, considerando que su carácter racional recae no tanto en la conducta ni en los beneficios que ésta acarrea para el organismo, sino en el proceso de selección natural que le dio origen. Esto es así dado que no puede hablarse de racionalidad si no hay un espacio de posibilidades de entre las cuales la acción haya sido elegida. Comportamientos desencadenados automáticamente ante la presencia de un determinado estímulo no pueden ser considerados como racionales a nivel individual; sin embargo, el hecho de que hayan sido preservados por la selección natural de entre una gama de *alternativas* disponibles a lo largo del tiempo evolutivo lleva a Bermúdez (2003) a atribuirles un *carácter racional derivativo*.

En el otro extremo Bermúdez (2003) coloca el “nivel 2” de racionalidad, que delimita los comportamientos soportados por *procesos* racionales y cuyo aspecto definitorio es la toma de decisiones:

Pero decir que una acción es llevada a cabo con base en una sensibilidad a las consecuencias conlleva mucho más que el sólo hecho de que sea realizada a causa de sus consecuencias. Implica que el agente ha hecho una evaluación de esas consecuencias sobre la base de una creencia acerca del probable resultado de una acción (y, más probablemente, de una comparación de este resultado con los probables resultados de otros posibles cursos de acción). En otras palabras, la toma de decisiones sólo tiene lugar en un nivel donde se dispone de creencias instrumentales (p. 124).

Tradicionalmente se ha pensado que todo comportamiento puede ser incluido en una de estas dos grandes categorías. De esta forma, tendríamos conductas que son directamente desencadenadas por estímulos ambientales particulares y otras que son consecuencia de un proceso sistemático de deliberación que opera sobre información representada internamente. No obstante, y como ya lo he comentado, esta dicotomía nos lleva a la embarazosa consecuencia de que, toda acción que no sea explicable como resultado directo de un estímulo desencadenador, habría de serlo a partir un proceso teórico de razonamiento y viceversa (Fodor, 1975; Mitchell, De Houwer y Lovibond, 2008).

Lo dicho hasta aquí hace evidente que entre el “nivel 0” y el “nivel 2” de racionalidad existe un vasto terreno de evidencia empírica que demanda una explicación alternativa. Con esto en mente, Bermúdez (2003) postula la existencia de un nivel de racionalidad adicional, el “nivel 1”, que da cuenta de acciones que ni son automáticamente provocadas por el ambiente, ni son consecuencia de un proceso explícito de razonamiento. Para ello, echa mano de la teoría de los *affordances* desarrollada por James J. Gibson (1986) en su obra *The Ecological Approach To Visual Perception*. Para Gibson (1986, p. 127): “los *affordances* del ambiente son aquello que éste *ofrece* al animal, con lo que el ambiente lo *provee* o *equipa*.” En palabras de Bermúdez (2003):

La idea básica detrás de la teoría de los *affordances* de Gibson es que el ambiente no es percibido en términos neutrales. Lo que es percibido son las posibilidades que el medio proporciona (*affords*) para la acción

y la reacción [...] La percepción de affordances es una percepción de las propiedades instrumentales del ambiente (p. 47).

Un aspecto muy importante de la propuesta de Gibson (1986) es que nos obliga a abandonar nuestra concepción puramente contemplativa de la percepción a cambio de una más dinámica y orientada sobre la acción. Como comenta Bermúdez (2003):

Si [Gibson] está en lo correcto, entonces podemos observar como una conducta dada puede ser seleccionada de un rango de alternativas en una forma que no involucra un proceso de toma de decisiones. La comparación de los affordances no requiere de un proceso de toma de decisiones. No obstante, es evaluable de acuerdo a un criterio de racionalidad. Llamaré a esta la esfera del nivel 1 de racionalidad (p. 121, corchetes añadidos).

La diferencia básica entre el “nivel 1” y el “nivel 2” es que, mientras que en el segundo caso el componente instrumental en el esquema de explicación psicológica toma la forma de una creencia instrumental separada, en el primer éste es parte integral del contenido de la percepción. La introducción del “nivel 1” de adscripción racional tiene implicaciones muy profundas para nuestras prácticas explicativas en ciencias cognitivas. De entrada, rompe la dicotomía que contraponía procesos automáticos vs procesos racionales. No se trata pues ya sólo de establecer si una conducta es o no racional (*i.e.*, susceptible de explicación psicológica), sino de determinar *en qué sentido* una conducta es racional. Un reto adicional consiste

en dilucidar la forma en que los diferentes niveles de adscripción racional interactúan y se complementan para permitirnos movernos en el mundo.

Resumiendo, caracterizar los procesos cognitivos como procesos computacionales no nos ofrece teoría alguna sobre la racionalidad (por ejemplo, podría programarse una computadora para que actuara “irracionalmente”), por lo que la especificación de la arquitectura simbólica de la mente no guarda una relación transparente con el carácter sistemático de nuestro pensamiento (como lo pretendía Fodor y Pylyshyn, 1988). No obstante, es un hecho que la teoría computacional de la mente resultó un impulso muy fuerte para la legitimización de una conceptualización de la racionalidad que se remonta, cuando menos, hasta Descartes: la idea de que pensar racionalmente es equivalente a pensar a partir de las reglas de la lógica y la matemática. No obstante, a lo largo de este capítulo argumenté que definir y evaluar la racionalidad de nuestras acciones no es para nada una tarea sencilla. Más allá del hecho de que gran parte de nuestra conducta inteligente podría no estar soportada en *procesos* racionales, investigaciones como la de Antonio Damasio (1994/2006) o la de Gerd Gigerenzer y sus colaboradores (2004; Gigerenzer y Brighton, 2008; Gigerenzer y Gaissmaier, 2011; Todd y Gigerenzer, 2000) apuntan fuertemente a que la noción misma de racionalidad omnisciente bien podría no ser más que un mito.

CAPÍTULO SEIS
¿“ADICTOS” A LA INFORMACIÓN? LA RACIONALIDAD COMO
FENÓMENO NATURAL

Ninety percent of life is just being there.

Woody Allen

Hasta este punto he intentado mostrar que hay al menos una forma en la que es posible tener conducta propiamente racional (cuando es evaluada a partir de sus consecuencias) sin un proceso explícito de deliberación. Es precisamente esto lo que nos ofrece la teoría de los *affordances* de Gibson (1986) y los tres niveles de adscripción racional propuestos por Bermúdez (2003). No obstante, falta por resolver el problema de la continuidad entre nuestros procesos de razonamiento y aquellos de las especies carentes de lenguaje.

Una primera consideración que debería saltar a la vista es lo injustificado que resulta considerar que exista una caracterización unitaria del “pensamiento animal” contrapuesta a la cognición humana. Los estudios de cognición comparada sólo tienen sentido si parten del contraste de las competencias cognitivas de *una* especie con las de *otra*, y no de la pretensión

de que es posible comparar la cognición de *una* especie con *la* cognición de *todas las demás*. Este hecho nos obliga a redefinir nuestro objetivo como la búsqueda de una configuración de capacidades y sistemas cognitivos relativamente constante a lo largo de escalar filogenético (presente en especies cuya cognición diferiría, en mayor o menor grado, entre una y otra) y la determinación de la forma en que nuestra psicología se relaciona con este esquema general.

El capítulo final de esta tesis tiene como objetivo explorar hasta qué punto podemos ofrecer una caracterización naturalista de la racionalidad que nos permita dar cuenta del problema de la continuidad cognitiva al tiempo que haga justicia a la complejidad y sofisticación de los sistemas humanos de conocimiento. Si bien acepto que aún es necesaria mucha investigación y discusión para aclarar muchos de los aspectos particulares que aquí sólo alcanzo a introducir, también considero que en los últimos años ha comenzado a emerger una imagen de la cognición que se presenta como muy fructífera para intentar comprender las intrincadas relaciones existentes entre lenguaje, cognición y comportamiento. Así pues, aún con las muchas y profundas limitaciones inherentes a un trabajo como el que aquí presento, espero que esta reconstrucción pueda contribuir, aunque sea un poco, a revitalizar la discusión entre los interesados en el estudio de la cognición y la racionalidad.

6.1. ¿CONTINUIDAD O DISCONTINUIDAD COGNITIVA? LAS CARAS DEL PROBLEMA

A grandes rasgos, persisten dos diferentes maneras de encarar el problema de la continuidad. La primera supone que la organización global de la cognición humana, aun siendo consecuencia de la evolución biológica, difiere esencialmente del esquema organizativo de la cognición de otras especies —por ejemplo, por el surgimiento de alguna capacidad que hubiera llevado a una reestructuración radical del sistema, como lo es el lenguaje para Vygotsky (1934/1986). La segunda aproximación presume que la cognición humana se organiza en torno al mismo núcleo de capacidades que la de otras especies y que las diferencias pueden ser trazadas en aspectos muy concretos del funcionamiento mental (Bermúdez, 2003; Mercier y Sperber, 2011). Lo dicho hasta ahora ha de hacer evidente que me inclino por la segunda opción; sin embargo, resulta conveniente considerar a detalle la primera para apreciar algunas de sus limitaciones.

Penn, Holyoak y Povinelli (2008) sostienen que:

Aunque los animales humanos y los no-humanos comparten muchos mecanismos cognitivos similares, nuestra *hipótesis de la reinterpretación relacional* (RR) sostiene que sólo los animales humanos poseen los procesos representacionales necesarios para reinterpretar sistemáticamente las relaciones perceptuales de primer orden en estructuras de orden superior, gobernadas por reglas, como aquellas postuladas en un *sistema físico de símbolos* [*physical symbol system*, PSS, por sus siglas en inglés] (p. 111, corchetes añadidos).

Un sistema físico de símbolos (PSS) parte del supuesto de que: “las representaciones mentales están compuesta de tokens simbólicos discretos, que pueden ser combinados en representaciones complejas formando relaciones sintácticamente estructuradas de varios tipos” (p. 123). ¡Penn, Holyoak y Povinelli (2008) argumentan que, después de todo, Putnam, Fodor y Pylyshyn no estaban del todo equivocados!: aunque bastante inadecuada como caracterización general de la cognición, tal parece que la hipótesis PSS describe, en general, de forma adecuada el proceso humano de pensamiento.

No voy a negar aquí que los miembros de nuestra especie puedan aproximar la manipulación explícita de símbolos supuesta por la hipótesis PSS. Por ejemplo, en este momento, al tiempo que escribo en mi ordenador, procuro vincular de forma coherente las proposiciones que estructuran mi discurso (aunque seguro son menos éxito del que yo esperaría), de tal forma que resulte tanto comprensible como atractivo y convincente para los posibles lectores. Lo mismo se puede decir de los casos en los que me veo enfrentado a resolver un problema algebraico o a desarrollar una prueba lógica. Pero el que haya sacado 10 en mi examen de lógica no demuestra que emplee el cálculo de predicados para resolver los problemas a los que me enfrento *más allá del contexto específico* en que se me demanda hacer uso de ese conocimiento.

Esta crítica puede ser extendida a la totalidad de la investigación en ciencias cognitivas. Evidenciar que un determinado organismo es capaz de resolver la tarea definida por un determinado diseño experimental *no*

nos dice nada de la forma en que, de hecho, aprovecha esa capacidad en su interacción cotidiana con el mundo. Dicho de otra forma, demostrar experimentalmente la posesión de una competencia cognitiva es sólo la mitad del trabajo de un investigador; la otra parte, tan importante como la primera (aunque casi siempre pasada por alto), consiste en revelar la manera en que ésta entra en juego en los contextos ecológicamente relevantes *para el organismo.*

Pienso que algunas de las más acaloradas discusiones en el terreno de la cognición comparada se alimentan de la sistemática costumbre de reducir el estudio del pensamiento animal a intentar demostrar que los miembros de especies diferentes de la nuestra son capaces de resolver tareas que sólo considerábamos accesibles a nosotros. Un ejemplo particularmente claro de este hecho es el estudio de las habilidades lingüísticas en primates no-humanos (véase, por ejemplo, Savage-Rumbaugh, Shanker & Taylor, 1998; Premack y Premack, 1988). El problema surge cuando, a partir del hecho de que *un* bonobo o *un* chimpancé particular alcancen cierto grado de maestría en el uso de un sistema simbólico de comunicación, se pretende argumentar que los mecanismos cognitivos que sustentan el lenguaje humano ya están *plenamente desarrollados* en los bonobos y los chimpancés (pasando por alto que, de entrada, estamos muy lejos de comprender cuáles son los mecanismos cognitivos que subyacen al aprendizaje y uso del lenguaje *en nuestra especie*).

No obstante, como lo comenta Michael Tomasello (2005; 2008) este tipo de generalizaciones resultan por demás injustificadas: de ser cierto que

otras especies poseen mecanismos “lingüísticos” *análogos* a los nuestros, ¿por qué no los usan de forma análoga a como lo hacemos nosotros? O, con otras palabras, no se puede defender la existencia de sistemas cognitivos equivalentes que lleven a configuraciones conductuales radicalmente distintas. Pretender que es posible extraer conclusiones generales sobre la organización cognitiva de un organismo a partir de su actuar ante un paradigma experimental aislado no puede más que oscurecer nuestra comprensión de los fenómenos psicológicos: *es sólo a partir de nuestro conocimiento sobre los patrones globales de actividad de un organismo que podemos dar sentido a su desempeño ante una determinada tarea* (y no viceversa).

Así pues, la diferencia básica con la propuesta de Penn, Holyoak y Povinelli (2008) no radica en determinar si los seres humanos somos capaces de *reflexionar sistemáticamente sobre la estructura del mundo* (¡claro que lo somos!), sino en determinar las consecuencias que este hecho (pienso, innegable) tiene para *la forma en que nos movemos en el mundo*. Dicho de otro modo, la cuestión está en determinar si nuestra capacidad de “reinterpretar sistemáticamente las relaciones perceptuales de primer orden en estructuras de orden superior, gobernadas por reglas” (p. 111) está restringida a aspectos específicos de nuestra actividad (Bermúdez, 2003; 2003a) o si es transversal a toda (o a gran parte) de nuestra cognición (Penn, Holyoak y Povinelli, 2008).

Penn, Holyoak y Povinelli (2008) no están del todo injustificados para defender que existan procesos de *reinterpretación relacional* transversales a diferentes dominios de nuestra cognición. Ellos mismos han mostrado

de manera convincente que sólo los humanos somos capaces de representarnos el mundo físico (Penn y Povinelli, 2007) y el mundo social (Penn y Povinelli, 2007a) como gobernados por reglas abstractas, más allá de lo inmediatamente perceptible, y de razonar sistemáticamente a partir de estas reglas. No obstante, su argumento depende de que se demuestre que los procesos de razonamiento que las personas usan cuando colaboran en un *experimento psicológico* (cuyo planteamiento y resolución casi siempre son dependientes del lenguaje) son los mismos que guían su actuar cuando se enfrentan a problemas de la “vida real”, que poseen la estructura que el psicólogo *intentó* capturar en su diseño experimental. ¿Existe evidencia empírica que nos permita evaluar esta posibilidad? Aunque aún es necesaria mucha la investigación en esta dirección, contamos con algunos resultados sugerentes que nos llevan a rechazar la conclusión defendida por Penn, Holyoak y Povinelli (2008).

6.2. CUANDO EL MÁS ES MENOS O ¡NO LO PIENSES DEMASIADO!

Es claro que hay muchos que aún estarán dispuestos a dar la última batalla para defender nuestro estatus como animales plenamente racionales, y que mejor forma de hacerlo que regresar a Descartes y separar el nivel teórico del nivel práctico de nuestra actividad. La estrategia podría ser como sigue: bien, podemos aceptar que las exigencias de la vida cotidiana muchas veces nos obligan a resolver los problemas a partir de estrategias heurísticas como las propuestas por Gigerenzer (2004; Gigerenzer y Brighton,

2008; Gigerenzer y Gaissmaier, 2011; Todd y Gigerenzer, 2000) dado que siempre es mejor tomar una decisión a partir de información incompleta que dejar pasar una oportunidad por no decidir a tiempo. Pero, sin duda alguna, siempre sería mejor actuar a partir de una consideración detallada de todas las posibilidades y de una evaluación sistemática de sus consecuencias. El ejercicio íntegro de nuestra capacidad de razonamiento nos llevaría a tomar mejores decisiones y, con ello, a ser más felices, por lo que habríamos de aprovechar todo nuestro potencial cognitivo en las situaciones que así no lo permitan.

Sin embargo, esta posibilidad resulta contradicha por la evidencia empírica de la que actualmente disponemos. Una muestra de esto es el llamado efecto de “más es menos”: más tiempo, más información y más deliberación, ¿no nos llevan necesariamente a tomar mejores decisiones! Por ejemplo, está ampliamente documentado que aquellas personas que son llevadas a considerar con detalle las posibilidades y consecuencias de sus elecciones (los “maximizadores”) —ya sea por rasgos inherentes a su personalidad, ya sea por exigencias impuestas por el diseño experimental—, eligen distinto y suelen reportar mayor arrepentimiento y menores índices de satisfacción que aquellas que realizan una búsqueda limitada de opciones y eligen la primera “suficientemente buena” (los “satisfactores”) (Dijksterhuis y Olden, 2006; Gigerenzer, 2007; Schwartz, 2004; Wilson y Schooler, 1991; Wilson, Lisle, Schooler, Hodges, Klaaren y LaFleur, 1993).

Un efecto similar ha sido reportado en la elección de conductas motoras. Beilock y sus colaboradores (2002; 2004 citados en Gigerenzer, 2007)

mostraron que golfistas expertos a quienes (1) se les restringió su tiempo de golpeo a tres segundos o (2) se les pidió que tiraran al tiempo que realizaban una tarea de distracción, llevaron a cabo mejores golpes que en las condiciones donde no tenían restricción de tiempo o se les pedía que concentraran toda su atención en el tiro. De forma importante, los efectos opuestos fueron observados en golfistas novatos: ellos tiraban mejor en condiciones sin restricción de tiempo y cuando se les pedía que se concentraran en su golpe (véase Gigerenzer, 2007 y Schwartz, 2004 para más ejemplos del efecto “más es menos” y para una discusión más profunda sobre sus implicaciones para una teoría de la elección).

Por el momento, me gustaría considerar con detalle otro interesante efecto que desmiente la idea de que nos guiemos por la búsqueda de la maximización, incluso en las situaciones que nos permitan una reflexión detallada del problema. El hecho que quiero destacar es que muchas veces nuestra elección consiste simplemente en ¡no tomar una decisión!, incluso en contextos que resultan muy relevantes desde el punto de vista social o personal.

Imaginemos un problema al que todos nos podemos enfrentar en nuestra vida cotidiana: decidir entre participar o no en un programa de donación de órganos. Aunque en un primer momento resulta oscuro determinar cómo es que un evento que sólo puede tener lugar después de que ya no estemos aquí nos permitiría aumentar nuestras utilidades (¿cómo es que algo puede beneficiarme —o perjudicarme— después de muerto?!), al considerar las implicaciones sociales de una elección de esta naturaleza es claro

que aspectos como el reconocimiento o el estatus social deben jugar un papel muy importante (también se podría pensar la situación en términos de *fitness* inclusivo; es decir, a partir de los beneficio o perjuicios que mi decisión podría tener sobre aquellos que comparte mi material genético).

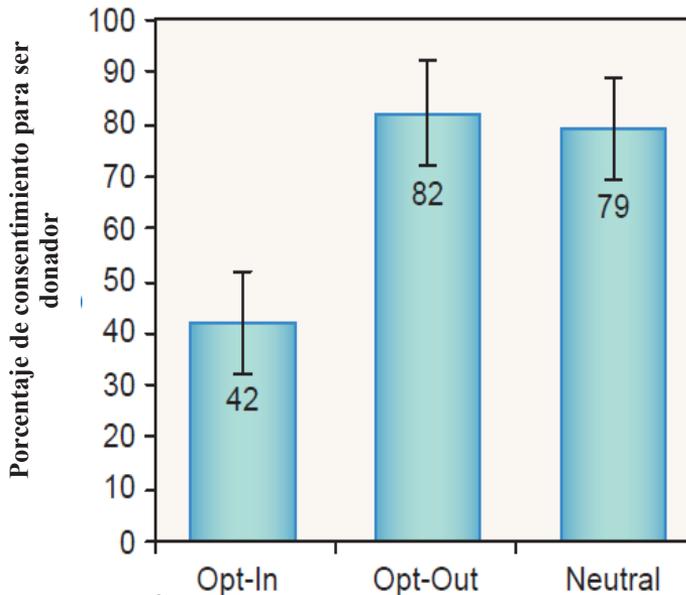


Figura 1: Índices efectivos de consentimiento como una función de las opciones por defecto: Opt-In: no ser donador por defecto; Opt-Out: ser donador por defecto; Neutral: elección sin default. Adaptado de Johnson y Goldstein (2003).

¿Cómo es que la mayor parte de la gente resuelve un problema de este tipo? Eric J. Johnson y Daniel Goldstein (2003) ofrecen una respuesta provocativa: eligiendo no elegir. Estos autores exploran con particular detalle las consecuencias que las alternativas *por default* tienen sobre este tipo de decisiones. Johnson y Goldstein (2003) preguntaron a un grupo de personas si estaría dispuestos a donar sus órganos en tres condiciones distintas: (1) **Opt-In**: donde por default se les consideraba como no-donadores y podrían

cambiar a la opción de ser donadores con un clic sobre un formato electrónico presentado por medio de un ordenador; (2) **Opt-Out**: contraria a la opción anterior, por lo que se presuponían como donadores al menos me marcaran la opción contraria; y (3) **Neutral**: donde no se partía de ningún default y ellos tenían que marcar la opción de ser donadores o no serlo.

En la figura 1 se presentan los porcentajes de consentimiento para cada una de las condiciones. Es importante destacar que la marcada diferencia entre las condiciones Opt-In y Opt-Out es por completo inexplicable desde la teoría económica clásica o desde la aproximación logicista al estudio de la racionalidad, dado que ambos problemas son *formalmente equivalentes* (i.e., representan el mismo problema lógico).*

Tampoco pueden ser explicables por medio de efecto de *framing* propuesto por Kahneman y Tversky (1984) —que indica que replantear el

* No voy a considerar con detalle aquí la condición “Neutral”, dado que considero que la información de la que disponemos no nos marca una línea lo suficientemente sólida para desarrollar una interpretación en alguna dirección en particular. Con todo, me gustaría avanzar la siguiente explicación: Los resultados parecen indicar que los sujetos consideraron equivalentes las condiciones “Opt-Out” y “Neutral”, como lo revela el hecho de que los índices de consentimiento de ambas son prácticamente idénticos. Esta posibilidad no carece por completo de plausibilidad si consideramos que resulta difícil defender la neutralidad de una pregunta abierta sobre un tema tan vigente y socialmente relevante como lo es la donación de órganos: podría decirse que cuando alguien nos cuestiona “¿estaría usted dispuesto a participar en un programa de donación de órganos?”, implícito en la pregunta, ya hay un “por default” afirmativo... si no tuviera la expectativa de que contestáramos que “sí”, ¿por qué habría de preguntarnos en primera instancia?. La propuesta de Mercier y Sperber (2011) ofrece relativo apoyo a esta interpretación: existe una marcada tendencia a ajustar nuestras respuestas a lo que consideramos que es socialmente esperable, aunque no siempre refleje fielmente nuestras opiniones y preferencias sobre un asunto (esta consideración también suele figurar entre las explicaciones del efecto de “más es menos”). El contraste desarrollado enseguida entre los resultados del experimento “virtual” de Johnson y Goldstein (2003) y los índices reales de donación en diferentes países también parece estar apoyando esta interpretación. En última instancia, tanto la condición “Opt-Out” como la “Neutral” nos estarían exigiendo el desarrollo de un razonamiento explícito que nos permitiera justificar nuestra negativa a ser donadores, algo que podría demandarnos muchos recursos cognitivos y emocionales.

mismo problema en términos de pérdidas o ganancias tiene consecuencias muy marcadas sobre la manera en que las personas eligen—, como lo muestra el hecho de que más del 80 % de los participantes a los que se les enfatizaron las pérdidas (en este caso, la pérdida de sus órganos tras su muerte) persistieran en esta elección y que incluso el 40 % de los participantes en la condición Opt-In viraran hacia ella. Sin embargo, más investigación es necesaria para descartar otras interpretaciones de la tarea consistentes con este marco teórico y para dilucidar con mayor precisión la manera en que la semántica en la que se plantea la elección podría afectar el comportamiento (por ejemplo, planteando el problema a partir del número de personas que pueden vivir/morir como consecuencia de nuestra decisión).

Incluso la teoría del “marcador somático” de Damasio (1994/2006) es incapaz de dar cuenta de estos resultados ya que, si sólo los dos posibles cursos de acción (ser donador vs no serlo) están “marcados” positiva o negativamente en virtud de sus consecuencias, no habría razón para suponer que la inclusión de una elección por default debiera hacer una diferencia en los índices de consentimiento entre una condición y la otra. Pienso que efectivamente hay una reacción emocional detrás de los datos reportados por Johnson y Goldstein (2003), pero que ésta no opera sobre los cursos de acción abiertos por el problema, sino sobre el contexto mismo de la deliberación.

De este modo, al considerar la posibilidad de donar nuestros órganos, nos enfrentamos a una situación de elección marcada por la consideración de nuestra propia muerte, las consecuencias que la decisión podría tener

sobre la vida de otras personas y los *costos* de no tomar la decisión *correcta* (con todo lo que este “correcta” pueda significar en el nivel personal, social e, incluso, teórico), ente otros factores de carácter más idiosincrático (por ejemplo, nuestras convicciones religiosas, nuestras opiniones sobre el sistema de salud, el hecho de que alguna persona cercana haya recibido un trasplante y un largo etcétera). En su conjunto, estos elementos crean una situación que, a todas luces, podría ser descrita como propicia para la búsqueda de una *elección óptima* desde la teoría económica clásica.

No obstante, estos datos parecen estar indicando: (1) o la inclusión de una opción por default determina el curso del razonamiento (observación que ya es contraria a cualquier apelación un proceso óptimo de decisión) o (2) hay un gran número de personas que simplemente no considera a detalle la situación y prefiere quedarse con la alternativa marcada de antemano. Se podría argumentar que no tomar una decisión también conlleva una elección (¡y claro que es así!), pero de lo que aquí se trata no es de discutir si elegimos o no elegimos, sino de determinar cuál es la mejor forma de describir nuestros procesos cotidianos de toma de decisiones.

Considero hasta un lugar común la afirmación según la cual hay ciertos dilemas (como reflexionar sobre participar en un programa de donación de órganos) cuya consideración detallada resultaría tan costosa en términos cognitivos y/o emocionales (aunque dispongamos de mucho tiempo para repasar las opciones) que optar por suspender el juicio resultaría, en muchos sentidos, la mejor elección que podríamos tomar. Sin embargo, la única forma inteligible en que esta idea podría apoyar una teoría de la

racionalidad óptima (entendida como *proceso*) es afirmando que la deliberación que nos llevó a suspender el juicio ¡es tan compleja como el propio juicio que supuestamente nos ayudo a evitar!

Dado el prominente papel que autores como Damasio (1994/2006) otorgan a nuestras emociones en el razonamiento, considero que es más natural afirmar que éstas juegan un papel muy importante no sólo en la consideración de las alternativas abiertas por un problema planteado (y analizado) a nivel proposicional, sino también en la *percepción* de nuestro entorno inmediato y de las acciones que podemos operar sobre él. Este proceso sería más “básico” que nuestra capacidad de deliberación consciente en la medida en que se piense que nuestra facultad de evaluar la relación interactiva que establecemos con el mundo es un prerequisite (acaso evolutivo) sin el cual nos sería imposible reflexionar sobre nuestras conductas pasadas y futuras. En este sentido, resulta posible afirmar que buena parte de nuestra actividad cotidiana está mediada por mecanismos similares a los *affordances* gibsonianos (Gibson, 1986), siendo así explicable en términos de “nivel 1” de racionalidad propuesto por Bermúdez (2003), y que nuestra capacidad de pensar el mundo en términos abstractos complementa, pero no sustituye, los sistemas de percepción-acción presentes en otros animales. No obstante, antes de pasar a revisar con detalle esta posibilidad, me gustaría regresar a lo reportado por Johnson y Goldstein (2003).

He enfatizado la importancia de distinguir entre el comportamiento emitido en nuestra interacción cotidiana con el mundo y aquel que se restringe a los ambientes experimentales —y, en algún sentido, anormales—

de la evaluación psicológica, así como la imposibilidad de obtener conclusiones generales a partir de la conducta emitida ante un único contexto de observación. De este modo, alguien podría estar inclinado a reprocharme que estoy pasando por alto mi propia aseveración y que pretendo extraer “principios últimos” sobre la cognición sólo soportados por las respuestas de un grupo restringido de sujetos ante el cuestionamiento de si participarían en un programa de donación de órganos. Por ejemplo, se podría argumentar que, efectivamente, muchas personas conservan la opción presupuesta en las condiciones Opt-In y Opt-Out del experimento de Johnson y Goldstein (2003), pero que, en un contexto con mayor validez ecológica —donde nuestra decisión en verdad tuviera consecuencias— habría de esperarse que el efecto de incluir una opción por default desapareciera.

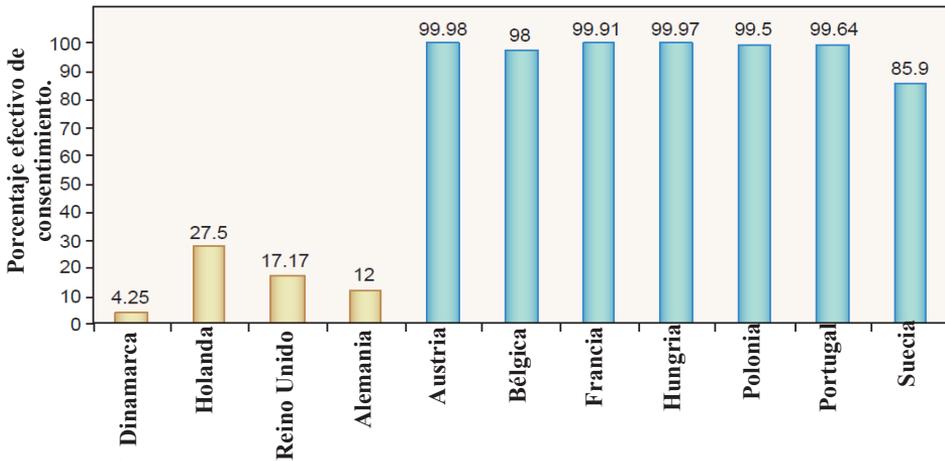


Figura 2: Índices efectivos de consentimiento por país. Consentimiento explícito (Opt-In: dorado); Consentimiento presupuesto (Opt-Out: Azul). Adaptado de Johnson y Goldstein (2003).

Con todo, e incluso aceptando el carácter tentativo de muchas de las ideas formuladas a lo largo de este trabajo, considero que lo que aquí se

ha dicho tiene un grado bastante aceptable de plausibilidad, aún sin ser de ninguna forma conclusivo (como no puede pretender serlo ninguna afirmación que se precie de ser científica). De entrada, y como he intentado mostrar, existe una amplia evidencia empírica detrás del efecto de “más es menos”, que ya por sí misma sustenta el argumento de que gran parte de nuestro razonamiento no opera a partir de la consideración detallada de todos los elementos relevantes para tomar una decisión y, lo que es más, que un *exceso* de tiempo, deliberación o información no necesariamente se traduce en una mejor elección.

Por otra parte, Johnson y Goldstein (2003) sustentan sus afirmaciones no sólo en los resultados arrojados por su “experimento virtual”, sino también en los índices *reales* de consentimiento en diferentes países de la Unión Europea cuyas legislaciones consideran el consentimiento presupuesto o el consentimiento explícito en lo que tiene que ver con la donación de órganos. Estos datos, ¿soportan la predicción del teórico de la elección óptima según la cual el efecto de la inclusión de una opción por default desaparece al aumentar la validez ecológica del problema o apoyan la interpretación de que, ante contextos de decisión muy demandantes, muchas personas prefieren simplemente no elegir? En la figura 2 se muestran los índices *reales* de consentimiento en once diferentes países. En las barras claras (primeras cuatro de izquierda a derecha) se presentan los países cuyo marco legal exige que las personas hagan explícita sus decisión de ser donadores (Opt-In) y las barras más oscuras aquellos donde se presupone que un individuo es donador a menos que exprese su voluntad de no serlo (Opt-Out).

Ha de ser evidente que la diferencia entre los dos grupos de países es incluso más marcada que cuando se considera la situación sólo a nivel hipotético: el promedio de consentimiento de los países bajo la condición Opt-In está a penas por arriba del 15 %, mientras que en los países englobados en Opt-Out este índice ¡supera el 97 %! El caso de Holanda resulta muy interesante dado que los datos presentados corresponden con un momento en que se emprendió una agresiva campaña gubernamental encaminada a concientizar a la población sobre la importancia de la donación de órganos (campaña que deberíamos considerar exitosa aunque sólo haya elevado el índice de consentimiento a 27.5 %).**

Resumiendo, cada vez es más la evidencia empírica que sugiere que: (1) más tiempo, más información y más deliberación no necesariamente nos llevan a tomar “mejores decisiones” y que, inclusive, se puede traducir en mayores índices de insatisfacción y arrepentimiento; (2) aun en los contextos propicios para la búsqueda de decisiones óptimas, las personas suelen guiarse por estrategias de carácter más difuso que pueden llegar hasta la simple suspensión del juicio; (3) si en verdad queremos compren-

** Es importante aclarar que los índices de consentimiento no se traducen directamente en el número de donadores. Son muchos los factores que aquí entran en juego. Primero, las circunstancias en que muere una persona no siempre hacen posible el aprovechamiento de sus órganos. Otra consideración de central importancia es que, incluso en los países con consentimiento presupuesto, la decisión última de donar suele recaer sobre la familia, muchas veces dejando de lado la propia legislación (Abadie y Gay, 2006). Con todo, Abadie y Gay (2006) argumentan, aún teniendo en cuenta estas consideraciones, la donación cadavérica en los países con consentimiento presupuesto suele ser de 25 a 30 % superior que en los casos en que se requiere el consentimiento explícito del donador. Esto nos debería llevar a considerar con detalle las implicaciones el consentimiento presupuesto tiene sobre la elección de la familia y el proceder de los médicos, además de abrir campo bastante fértil para explorar la forma en que las emociones entran en juego en este tipo de contextos de decisión (con todas las consideraciones éticas que merece emprender una investigación de esta naturaleza).

der la dinámica de los fenómenos cognitivos, es necesario que tengamos siempre presente que el contexto en que llevamos a cabo nuestros experimentos no siempre refleja de forma fiel la estructura de los problemas ecológicamente relevantes para los organismos; y (4) resulta sugerente pensar que una buena parte de nuestra actividad está mediada, no tanto por la reconstrucción abstracta que podamos realizar de una situación en términos proposicionales, sino por la percepción directa de los problemas —donde nuestras reacciones emocionales habrían de tener un papel muy destacado— y nuestras posibilidades de acción sobre ellos. En la siguiente sección voy a considerar con más detalle esta última observación.

6.3. PERCEPCIÓN Y ACCIÓN: EL INCONSCIENTE COGNITIVO.

Como lo comenté más arriba, la cuestión de fondo en las discusiones sobre la racionalidad tiene que ver con la relación existente entre razonamiento y conducta. ¿Hasta qué punto se puede decir que nuestras razones *causan* nuestras acciones? ¿Todo razonamiento antecede a la conducta que racionaliza o pueden darse casos en que la conducta sea previa? ¿Puede haber procesos de razonamiento que no se materialicen en comportamiento manifiesto? ¿Y qué decir de la posibilidad de que no podamos dar razones para algunos de nuestros *actos racionales*? Para una acción dada, ¿la relación entre razones y conducta permanece constante o cambia a través del tiempo? Estas son sólo algunas de las preguntas a las que una teoría de la acción racional se enfrenta; no pretendo dar una respuesta pormenorizada a

cada una de ellas, sino sólo ofrecer una panorámica general de las implicaciones que la investigación actual en ciencia cognitiva tiene para este tipo de cuestionamientos.

Hasta este punto he mostrado que hay una parte de nuestro comportamiento que puede ser explicada sin invocar un genuino proceso de toma de decisión; es decir, sin apelar a ningún proceso de deliberación explícita en términos proposicionales —*i.e.* que, como en el caso de otros animales, podemos dar cuenta de muchas de nuestras acciones a partir de lo que Bermúdez (2003) llama el “nivel 1” (vs el “nivel 2” de racionalidad). En estos casos decimos que tenemos una *conciencia fenoménica* del problema, pero no necesariamente una *conciencia reflexiva* de él: somos conscientes de aquellos elementos relevantes en el entorno inmediato que nos hacen actuar —los percibimos—, pero no somos conscientes de que, de hecho, somos conscientes de ellos —no los pensamos.

En muchos sentidos, la *conciencia reflexiva* es uno de los grandes misterios del mundo natural. Como Dijksterhuis, Chartrand y Aarst (2007) comentan:

Aún nadie ha ofrecido una respuesta convincente a la cuestión de por qué tenemos conciencia, y nosotros tampoco lo hacemos aquí. Entender por qué tenemos conciencia y cómo la conciencia funciona es uno de los grandes retos —si no es que el mayor de ellos— que la ciencia aún tiene por superar. [...]

La percepción consciente es sólo una pequeña parte de la percepción como un todo. Lo que es más, no parecemos ser más conscientes de aquellas percepciones que afectan directamente nuestra conducta de

lo que lo somos de aquellas que no lo hacen. [...] Incluso es posible argumentar en la dirección contraria. Muchos han enfatizado que “percibimos para hacer” pero esto aplica más en general a la percepción (consciente e inconsciente) que a la percepción consciente en sí misma. La percepción consciente puede fácilmente ser separada de sus consecuencias conductuales, como cuando uno conscientemente disfruta de la vista desde lo alto de un rascacielos sin experimentar cambio comportamental alguno. Del mismo modo, si llegamos a adquirir acceso consciente a un estímulo que puede afectar nuestra conducta, frecuentemente lo hacemos sólo después de haber actuado sobre él. Primero retiramos la mano y sólo entonces somos conscientes del dolor provocado por el pan caliente. *En suma, la percepción inconsciente es para hacer, pero la percepción consciente frecuentemente no* (pp. 113-114, *cursivas añadidas*).

Como hace evidente esta cita, del hecho de que poseamos una conciencia reflexiva de la situación no se sigue que el control de nuestras acciones recaiga de manera exclusiva, o cuando menos prioritaria, sobre esta percepción consciente. Dos alternativas resultan igualmente plausibles, tantos desde el punto de vista conceptual como desde el empírico: (1) incluso en los casos en que nos formamos una representación explícita de la situación, nuestro comportamiento puede verse influenciado por elementos del contexto que queden fuera de ella; o (2) más que estar causadas por un proceso explícito de razonamiento, nuestras acciones podrían *ser la causa* de nuestra conciencia reflexiva de un evento —decimos entonces que nuestras razones *no explican*, sino que *justifican* nuestro comportamiento (Dijksterhuis, Chartrand y Aarst, 2007; Baumeister, Masicampo y Vohs, 2011).

Resulta natural afirmar que aun nuestra consciencia reflexiva depende de procesos inaccesibles a nuestra reflexión, por lo que, en algún sentido, todas y cada una de nuestras acciones descansarían sobre procesos de naturaleza inconsciente (Dijksterhuis, Chartrand y Aarst, 2007). Considero que esta afirmación es cierta pero trivial. Lo que resulta relevante es determinar si, *una vez que nos representamos explícitamente una situación*, esta representación ejerce un control efectivo y directo sobre nuestro comportamiento. El sólo hecho de demostrar la existencia de *cierto grado* de disociación funcional entre razonamiento y conducta ya nos obligaría a un replanteamiento radical de nuestro esquema tradicional de explicación psicológica. La evidencia empírica revisada a lo largo de este capítulo ya apunta fuertemente en esta dirección; no obstante, resultará útil profundizar en ella.

Consideremos el primero de dos escenarios que acabo de plantear. ¿Pueden estímulos más allá de nuestra percepción consciente *de una situación* impactar la forma en que nos desenvolvemos en ella? Es importante observar que lo que aquí está a discusión no es si somos o no reflexivamente conscientes de estos estímulos (aunque queda claro que deberíamos serlo a nivel fenoménico), sino sólo si elementos del entorno que no resultan relevantes *en nuestro proceso de razonamiento* podrían no obstante serlo en la determinación de nuestras acciones —lo cual demostraría, cuando menos, que la explicación de nuestra conducta no podría ser agotada por la verbalización de nuestras razones, contrario a lo defendido por autores como Davidson (1963/2001).

Consideremos la siguiente situación presentada por John A. Bargh (2005):

Paris, 1986: el Doctor Lhermitte acompaña a dos de sus pacientes a diferentes partes de la ciudad. Ambos han sufrido un golpe que dañó partes de su corteza prefrontal, áreas críticas para la planificación y el control de la acción. [...] Después de haber pasado una media hora en el departamento del profesor, él los invita a salir hacia el balcón, mencionando casualmente la palabra *museo*, y entonces les permite entrar de nuevo. De repente su conducta se vuelve diferente: ambos escrudían con gran detalle las pinturas y posters que cuelgan en la pared, así como los objetos comunes sobre las mesas, como si se trataran de obras de arte contemporáneo. Enseguida el hombre entra en el dormitorio, ve la cama, se desnuda y entra en ella. En poco tiempo está dormido. A través de estas y otras situaciones, ninguno de los pacientes es capaz de notar algo inusual o extraño en su conducta (p. 37).

Es claro que el comportamiento de estas personas está en gran medida *controlado* por la estructura del ambiente. Sus lesiones les han impedido separar su conducta de los *affordances* que éste les presenta: ahí donde hay una cama, hay también una situación propicia para dormir, ahí donde se activa su percepción de un museo, hay una gran oportunidad de disfrutar de las “obras de arte”. Elkhonon Goldberg (2002) hace ver que esta excesiva dependencia del ambiente para la emisión de comportamiento es uno de los rasgos más característicos de las personas que han sufrido daño en su corteza frontal; él mismo reporta casos de pacientes entrando y saliendo de las habitaciones de los hospitales “sólo porque las puertas estaban ahí”.

Uno de los aspectos más intrigantes de este tipo de casos es, como lo comenta Bargh (2005), que aquellos que los padecen no creen que haya algo anormal con su forma de actuar.

Se podría objetar, y con toda razón, que de las manifestaciones conductuales de un trastorno neurológico no se pueden extraer —o al menos no directamente— conclusiones sobre las personas sanas. No obstante, tomados con cuidado, los datos surgidos de estudio de personas con déficits neuropsicológicos pueden ofrecernos interesantes intuiciones sobre la organización de nuestro proceso cotidiano de pensamiento y permitirnos desarrollar aproximaciones experimentales que den luz, a su vez, del funcionamiento y la organización de diferentes mecanismos cognitivos —incluido su nivel neurofisiológico.

Dado el preponderante papel de la corteza prefrontal en la regulación y organización de la conducta, no debería de sorprendernos que aquellas personas con lesiones en esta parte del cerebro pierdan la capacidad de desligar sus acciones de los *affordances* ambientales. ¿Es posible observar, en algún grado, este fenómeno en personas sanas? En los últimos años ha habido una explosión en el número de voces que argumentan que no sólo es posible, sino que, de hecho, es muy común (Bargh, 2005; Dijksterhuis y Olden, 2006; Gigerenzer, 2007; Wilson y Schooler, 1991; Wilson, Lisle, Schooler, Hodges, Klaaren y LaFleur, 1993; véase Dijksterhuis, Chartrand y Aarst, 2007 para una revisión pormenorizada de los trabajos en esta área).

Entre los resultados que mayor apoyo ofrecen a esta conclusión está el ampliamente documentado efecto de *priming*. La idea básica es que es

posible alterar el comportamiento de la gente ante una situación determinada sólo activando su representación de estímulos que, podría pensarse, nada tienen que ver con su contexto de actividad. Bargh, Chen, y Burrows (1996, citado en Dijksterhuis, Chartrand y Aarst, 2007) fueron los primeros en reportar el efecto que la activación de una categoría tiene sobre la conducta motora. Estos autores invitaron a un grupo de personas a participar en un experimento en el que, supuestamente, se iba a evaluar sus dominios de la gramática inglesa. Sus participantes tenían que formar sentencias gramaticalmente correctas a partir de secuencias de palabras aleatoriamente ordenadas. A la mitad de ellos, se les expuso a palabras estrechamente vinculadas a la categoría “anciano” en la sociedad norteamericana (por ejemplo, gris, bingo, bastón, Florida), mientras que la otra mitad completó la misma tarea pero con palabras que no guardaban una relación semántica transparente. Al final de todos los ensayos se dio las gracias a los participantes y se les dijo que el experimento había terminado. No obstante, se tomó el tiempo que les tomaba llegar al elevador más cercano. Aquellas personas que fueron expuestas a la categoría “anciano” ¡resultaron significativamente más lentos que aquellos en el grupo control!

Desde la demostración pionera de Bargh, Chen, y Burrows (1996, citado en Dijksterhuis, Chartrand y Aarst, 2007) muchos otros autores han corroborado que la activación de una categoría semántica lleva a las personas a actuar en concordancia con ella, aun cuando la tarea en que son evaluados sea totalmente independiente —por ejemplo, Nelson y Norton (2005, citado en Dijksterhuis, Chartrand y Aarst, 2007) mostraron que la

activación de la representación de superhéroes aumentó el grado de amabilidad de las personas.

De forma importante, tales efectos tienen lugar al margen de cualquier conocimiento explícito de esta relación por parte del participante y aun cuando la presentación de los estímulos se realice de forma subliminal (Dijksterhuis, Chartrand y Aarst, 2007). Se podría argumentar que tales procesos requieren mediación lingüística; sin embargo, encuentro por demás oscura la afirmación de que podamos operar a partir de representaciones lingüísticamente estructuradas que, no obstante, ¡nos sean verbalmente inexpresables! Tal afirmación sólo tendría sentido si sostuviéramos, con Fodor (1975), que todos los niveles de la cognición se organizan a partir de un lenguaje; considero que a estas alturas es insostenible tal afirmación.

El carácter perceptual y orientado a la acción de efecto de *priming* se ve apoyado por el hecho de que este puede ser elicitado a través de medios mucho más difusos que la activación directa de categorías semánticas. De este modo, Williams y Bargh (2008; véase también Bargh y Shalev, 2011) mostraron que personas que tuvieron contacto accidental con objetos cálidos (al sostener una taza de café, por ejemplo) evaluaron como más “cálida” su interacción con un desconocido. Ackerman, Nocera y Bargh (2010) han encontrado resultados análogos aplicables a otros contextos: la manipulación de objetos pesados hacen parecer a un candidato a un empleo como más importante, objetos ásperos nos llevan a percibir las interacciones sociales como más complicadas, y objetos duros aumentan nuestra rigidez ante las negociaciones.

Estos resultados nos llevan directamente a la segunda cuestión: ¿podría nuestro proceso de razonamiento ser más una *consecuencia* que la *causa* de nuestras acciones? Pensemos en el pan que nos quema la mano; es claro que nuestra experiencia consciente de la situación es sólo posterior a la emisión de una conducta adecuada ante ella (arrojar el pan y retirar la mano). Incluso se podría argumentar que en el caso de muchas de nuestras conductas reflejas no es necesaria representación explícita alguna: mis pupilas se contraen y se expanden ante las fluctuaciones de luz, empiezo a temblar ante un descenso en la temperatura, doy un salto y oriento mi mirada ante un ruido abrupto y todo, ¡por fortuna!, sin mediación consciente. No obstante, se podría objetar —de nuevo, de forma por demás adecuada— que este tipo de conductas ni siquiera requieren ser explicadas en términos psicológicos; ¿existe, pues, evidencia empírica que soporte la idea de que nuestro razonamiento, más que *causar* nuestras acciones “complejas”, las justifican? Una vez más la respuesta es afirmativa y, una vez más, la evidencia más clara proviene de la investigación neuropsicológica.

Consideremos el siguiente caso de una mujer quien quedó paralizada del brazo izquierdo tras un accidente cerebrovascular. William Hirstein (2005) lo describe de la siguiente forma:

El doctor regresa a la habitación, se aproxima a la mujer y la saluda. Cuando él le pregunta cómo se encuentra, ella reporta sentirse bien. “¿Sus dos brazos son igual de fuertes señora Esposito?” “Sí, ambos están bien”, responde ella. “¿Podría tocar mi nariz con su mano derecha?” Ella la levanta, de forma un tanto irregular, pero consigue tocar

la nariz del doctor. “¿Ahora podría tocar mi nariz con su mano izquierda?”, pregunta entonces el médico. La señora Esposito se detiene un momento, frota su hombro izquierdo y responde, “Oh, tengo artritis severa en mi hombro, sabe Doctor, es doloroso”. Ella no estaba mintiendo ni fingiendo, genuinamente creía que podía mover su brazo (p. 2).

Esta incapacidad de reconocer la propia enfermedad es conocida con el término genérico de *anosognosia* (Damasio, 1994/2006; Ramachandran y Blakeslee, 1998; Goldberg, 2002; Hirstein, 2005), en este caso concreto vinculada con una hemiplejía izquierda provocada por una afectación de la corteza somatosensorial derecha. Más allá del enigma que representa el hecho de que una persona pueda *ignorar* su propia condición médica (y aquí “ignorar” no es lo mismo que “negar” o “ser omiso”, sino que se acerca más a un *desconocimiento* de la situación), un aspecto muy interesante son los procesos de racionalización en los que los pacientes de embarcan para justificar su conducta. Como se ha dicho, las personas con anosognosia realmente *creen* que sus testimonios describen de forma adecuada las razones de su comportamiento, aunque éstos casi nunca invoquen a su historial médico o a las secuelas de sus enfermedades (y cuando lo hacen es sólo de forma tangencial, por ejemplo, diciendo que en el pasado tuvieron algún problema de salud pero que ahora están por completo sanos).

En los contextos neurológicos y psicológicos se suele decir que una persona *confabula* cuando recurre a complejas narraciones para dar cuenta de sus acciones, sin darse cuenta de la relación que éstas guardan con su

condición médica (o, lo que es más, sin siquiera reconocer que exista tal “condición médica”). Un aspecto importante de los procesos de confabulación es que tienen a preservar la consistencia interna de la narración, aunque ésta sea a todas luces contradicha por el estado de cosas en el mundo. Autores como V.S. Ramachandran (Ramachandran y Blakeslee, 1998) y Oliver Sacks (1988) han documentado con gran detalle el grado de elaboración que pueden alcanzar estas confabulaciones en diferentes trastornos neuropsicológicos; por ejemplo, Oliver Sacks (1988) reporta el caso de un hombre que tras despertar descubrió horrorizado que ¡los doctores había cosido un brazo que no era el suyo a su hombro izquierdo! (en realidad, había sufrido un accidente cerebrovascular que afectó el área encargada de procesar la información proveniente de su brazo en la corteza somatosensorial derecha, una situación muy similar a la de la mujer anteriormente descrita).

Además de la hemiplejía del lóbulo derecho, la confabulación está presente en otros desórdenes como el síndrome de Korsakoff (un trastorno de la memoria que lleva a quienes los padecen a formar complejas historias sobre sus actividades cotidianas que nada tienen que ver con la realidad; por ejemplo, dicen haber ido a cazar el fin de semana cuando en verdad estuvieron internos en el hospital), pacientes a los que se les ha practicado una disección radical del cuerpo calloso (una estructura compuesta de haces de axones que comunican los dos hemisferios cerebrales, impidiendo así que fluya gran parte de la información entre ambos. Es posible desencadenar una acción a partir de la presentación de una instrucción que sólo sea procesada por el hemisferio derecho; sin embargo, al ser cuestionadas

sobre el porqué de su conducta, las personas no tienen una consciencia explícita de la presentación de la instrucción y recurren a explicaciones que nada tienen que ver con ella) o la demencia de Alzheimer, entre otros trastornos (véase Hirstein, 2005, para un revisión detallada de estos desórdenes y su manifestación clínica).

No obstante, para que estos resultados resulten relevantes para la discusión que aquí he venido desarrollando, es necesario demostrar casos de genuina confabulación en nuestra vida cotidiana. Timothy D. Wilson (2002) ofrece algunos sugerentes resultados que apuntan en esta dirección. En uno de sus experimentos, se invitó a un grupo de mujeres a evaluar la calidad de cuatro diferentes marcas de medias de nylon. Cada uno de los pares fue etiquetado con las letras A, B, C o D y se dispusieron en una mesa de izquierda a derecha para que los participantes pudieran manipularlos. Los resultados mostraron que un 12 % prefirió las medias A, un 17 % las medias B, un 31 % las C y un 40 % las medias D. No obstante, ¡los cuatro pares de medias eran totalmente idénticos! Claramente, hay una tendencia a preferir las medias ubicadas más a la derecha, sin embargo, cuando se preguntó a las participantes sobre las razones de su elección, ninguna mencionó su posición sobre la mesa y se enfocaron en aspectos como la calidad del tejido, la textura, la elasticidad o la transparencia. Aún cuando se les dijo que los cuatro pares de medias eran idénticos, ¡las participantes insistieron en las razones de su elección!

La evidencia surgida de otras áreas de investigación también apoya, aunque de forma indirecta, la idea de que la mayor parte del tiempo real-

mente no somos conscientes de las causas de nuestras acciones (*i.e.*, de que las razones que podamos reportar sólo *después de actuar* no pueden ser vistas como la etiología de nuestra conducta). Pensemos de nuevo en la forma en que el efecto de “más es menos” impacta en la determinación de nuestras preferencias. Por ejemplo, Wilson y Schooler (1991) compararon el grado de acuerdo entre un grupo de catadores entrenados y otro de estudiantes al evaluar la calidad de diferentes marcas de mermelada de fresa.

Los estudiantes podían pertenecer a su vez a uno de dos diferentes grupos: (1) a los sujetos en la condición de “análisis de razones” se les pidió que contestaran un cuestionario en el que se les pedía expresar los criterios a partir de los cuales asignaron la calificación a cada una de las mermeladas; (2) los sujetos en el grupo control sólo tenían que puntuar cada marca sin ninguna justificación. Como habría de esperarse a partir de la revisión del efecto de “más es menos”, las evaluaciones entre los sujetos del grupo control y los catadores entrenados mostraron una alta correlación ($M = .55$), mientras que, en el caso de los estudiantes en la condición de “análisis de razones”, los juicios fueron marcadamente distintos ($M = .11$). La evaluación “razonada”, ¡prácticamente llevó a los participantes a invertir el orden de las mermeladas con respecto a los expertos!

El hecho de que una consideración detallada de la situación nos lleve a comportarnos, ya no sólo diferente a la forma en que lo hacen los expertos en un área, sino también a como lo haríamos nosotros mismos si no tuviéramos que entramparnos en un proceso de deliberación consciente, ofrece apoyo a la idea de que nuestros procesos de razonamiento *no siempre* guar-

dan una relación causal con nuestra conducta en la vida cotidiana.

Podría seguir enumerando investigaciones que echan por la borda la idea de que la verbalización de nuestras razones pueda ser vista como puerta directa a la causas de nuestras acciones, sin embargo, la lista sería demasiado larga. Con todo, me gustaría cerrar esta sección con la consideración de otro resultado empírico bastante interesante. En un trabajo que hoy ya podríamos considerar como clásico, Dutton y Aron (1974) exploraron hasta qué punto un estado general de activación podía ser interpretado como atracción sexual. Para ello, una (atractiva) científica confederada invitaba a un grupo de hombres a participar en una investigación sobre “los efectos de la exposición a atracciones escénicas sobre la expresión creativa” (p. 511). A cada uno de los sujetos se les presentó una imagen de una mujer joven cubriendo su rostro con una mano al tiempo que extendía la otra y se les pidió que escribieran una pequeña historia dramática sobre ella.

La investigadora abordaba a los hombres en una de dos condiciones distintas: (1) mientras atravesaban un puente sólido de 10 pies de alto sobre un pequeño riachuelo (grupo control), o (2) mientras atravesaban un puente colgante de 450 pies de largo que cruzaba un sistema de rápidos y formaciones rocosas a una altura de 230 pies (grupo experimental). Después de que los participantes terminaran de escribir su historia, la investigadora dio las gracias a cada uno de ellos y les dijo que, si querían, podían ponerse en contacto con ella para que los pusiera al tanto de los resultados de la investigación. Para ello, anotó a cada uno su número personal en una tarjeta y se despidió amablemente. Dutton y Aron (1974) conjeturaron que la condi-

ción experimental provocaría en los sujetos una activación fisiológica que se terminaría interpretando como atracción sexual, llevando a los hombres a establecer un contacto posterior con la investigadora. En línea con esta predicción, 9 de los 18 sujetos del grupo experimental la llamaron, mientras que esta relación fue de sólo 2 sobre 16 en el grupo control.

Aunque reveladores, estos resultados deben ser interpretados con cautela: bien podría darse cuenta de ellos apelando a diferencias intrínsecas entre los hombres de una y otra condición. Se podría argumentar, por ejemplo, que aquellos que cruzan el puente colgante podrían ser, de entrada, más intrépidos y tenderían más a la búsqueda de novedad que los hombres del grupo control (aunque esto tampoco equivale a decir que los últimos no lo pensarán cruzar). Para descartar esta posibilidad Dutton y Aron (1974) replicaron su estudio original (aunque esta vez fue otra mujer la que fungió como entrevistadora), pero ahora el grupo control estuvo conformado por varones que fueron abordados *al menos* 10 minutos después de haber atravesado el puente (razonando que este tiempo permitiría la disipación de la activación fisiológica provocada al cruzarlo). En línea con el experimento original, de los 20 sujetos que aceptaron el teléfono en el grupo experimental, 13 establecieron un contacto posterior, mientras que esta relación fue de sólo 7 sobre 23 en el grupo control.

Lamentablemente, el cambio de la entrevistadora no nos permite establecer una comparación directa entre los resultados de uno y otro experimento, aunque algunas cosas se pueden decir al respecto: (1) tal parece que, en general, la investigadora del experimento 2 resultó más atractiva

para los sujetos; y (2) los datos apuntan a que, efectivamente, existe una diferencia entre los hombres del grupo control de ambos experimentos. Finalmente, Dutton y Aron (1974) nunca justifican la idea de que 10 minutos son suficientes para “disipar” la activación fisiológica provocada al cruzar el puente, por lo que resulta imposible determinar si la diferencia observada entre los sujetos control de uno y otro experimento se debe a algún grado de excitación residual o a un rasgo de personalidad más estable. Me atrevo a avanzar la hipótesis de que el sólo hecho de estar en un contexto novedoso (como lo es el salir de vacaciones) ya provoca una diferencia con respecto a ser evaluado en un contexto familiar. Dado el sutil efecto observado en los sujetos controles, para evaluar esta posibilidad resultaría necesario aumentar el tamaño de la muestra y controlar elementos inherentes a la personalidad de los sujetos, por ejemplo, la búsqueda de novedad (hasta este punto, desconozco si se han realizado esfuerzos en este sentido).

Con todo, resulta claro que la relación entre razonamiento, emoción y comportamiento es mucho más compleja de lo que nuestros esquemas clásicos de explicación psicológica nos permiten apreciar. En concreto, los resultados de Dutton y Aron (1974) muestran que (1) algunas de nuestras racionalizaciones suelen estar encaminadas a justificar reacciones emocionales (y, en este sentido, son siempre posteriores a ellas) y que (2) podemos llegar a ser muy incompetentes a la hora de identificar las fuentes de nuestras reacciones fisiológicas. Un duro golpe más para la noción de razón omnisciente que prevalece en gran parte de la ciencia cognitiva contemporánea.

Este tipo de resultados han llevado a John A. Bargh (2005) a asegurar que “una disociación entre la conducta motora y el conocimiento consciente está ahora surgiendo como una propiedad estructural básica del cerebro humano” (p. 43). En este sentido, Timothy D. Wilson (2002) hace notar que:

Quizá la parte más radical [de este argumento] es que a pesar de la gran cantidad de información que la gente tiene, sus explicaciones sobre las causas de sus respuestas no son más adecuadas que las explicaciones que sobre ellas pudiera dar un completo extraño que viva en su cultura. ¿Cómo puede esto ser posible? ¿En realidad puede darse el caso en que un completo extraño, elegido aleatoriamente en el directorio telefónico, sepa tanto como nosotros sobre porqué respondemos en el modo en que lo hacemos? [...] Efectivamente, tenemos acceso a más información de nosotros mismos de la que podría disponer cualquier extraño. Sin embargo, puede que esta información no siempre nos lleve a inferir adecuadamente las causas de nuestras respuestas (pp. 108-109).

Los resultados revisados a lo largo de este capítulo deberían hacernos reflexionar sobre la conveniencia de seguir tomando las respuestas verbales (o escritas) que las personas puedan desarrollar ante la presentación de problemas h-i-p-o-t-é-t-i-c-o-s como una evaluación directa del su comportamiento ante las situaciones *reales* que las narrativas empleadas en nuestros diseños experimentales pretenden capturar. Aun reconociendo que, como lo han destacado autores como Paul L. Harris (2000/2005), los mecanismos fisiológico/cognitivos implicados en el procesamiento de escenarios ficticios son muy *similares* a los que entran en juego cuando tenemos una

experiencias directa de las situaciones, las investigaciones aquí revisadas hacen evidente que esta similitud no puede ser tomada como *equivalencia* y que, más que pretender explicar el comportamiento a partir de la verbalización que podamos hacer nuestras razones, *la relación entre razones y acciones es algo que debe ser explicado*.

De este modo, no sólo es ilícito pretender que se puede comparar de manera directa el desempeño de otras especies ante un problema *real* con lo que nosotros podamos decir de lo que *podríamos hacer* al enfrentarnos ante una situación análoga (una práctica bastante común en la investigación del comportamiento), sino que resulta igualmente arbitraria esta extrapolación al interior de nuestra propia organización conductual. En este sentido, Wilson (2002) continúa:

En primer lugar, parte de este conocimiento privilegiado no es tan adecuado como parece. Existe considerable evidencia de que la gente no es muy diestra en observar conscientemente la covariación entre sus respuestas y los antecedentes que la desencadenan. [...] Es muy común que haya muchos antecedentes de nuestras respuestas y nos resulta difícil determinar cuáles de ellos funcionan como causas. Dada esta dificultad, las creencias de la gente sobre covariación de manera frecuente son una función de sus teorías culturalmente transmitidas, más que una deducción basada en observaciones adecuada sobre su propia conducta. [...] Además, la gran cantidad de la información privilegiada de la que las personas disponen puede hacerles más difícil reconocer las causas de sus acciones de lo que podría resultar para un extraño que confiara en teorías culturales (pp. 109-110).

Sin duda alguna, mucha es la investigación que aún se necesita para esclarecer la intrincada relación entre razonamiento y conducta. Este trabajo resulta a todas luces insuficiente para siquiera intentar resolver esta cuestión, no obstante, me gustaría concluir con una somera revisión de una aproximación que, considero, se presenta como particularmente prometedora.

6.4. HACIA UNA NUEVA TEORÍA NATURALISTA DEL LA RACIONALIDAD HUMANA: EL RAZONAMIENTO Y LA VIDA EN SOCIEDAD

Hasta este punto, he sido por demás cauteloso a la hora de referirme de la relación existente entre razonamiento y acción. He echado mano de manera constante de expresiones atenuantes como “no siempre”, “cierto grado de disociación”, “algunas de nuestras conductas”, por mencionar sólo unas, a la hora de cuestionar la existencia de un vínculo causal directo entre razones y conducta. Esto ha sido así porque, considero, la evidencia disponible no soporta la idea de que nuestras razones *nunca* puedan ser tomadas como las causas nuestras respuestas.

Entiendo que hay tres criterios que nos permiten determinar cuándo es legítimo aseverar que una de nuestras acciones ha sido causada por nuestras razones: (1) la tarea no puede ser resuelta sin una reconstrucción previa, explícita y exhaustiva del problema; (2) la persona tiene acceso consciente a los pasos necesarios para transitar de una situación problema, A, a un estado de cosas deseable, B, o solución; y (3) la persona deliberadamente

implementa los pasos entre A y B en un orden preestablecido y es capaz de evaluar, al menos en principio, hasta qué punto el resultado se corresponde con la solución buscada. Como lo mencioné más arriba, instancias de este tipo de actividades son la solución de un problema matemático o el desarrollo de una prueba lógica, aunque también son buenos ejemplos el uso que podemos hacer de mapas para transitar por un espacio geográfico desconocido o el empleo de manuales para armar un mueble o preparar una nueva receta de cocina.

Empero, afirmar que en estos contextos nuestro razonamiento puede ser entendido como la causa de *algunas* de nuestras acciones no es lo mismo que decir que éstas dependan por completo de procesos de deliberación consciente. Como se mencionó más arriba, incluso nuestra consciencia reflexiva descansa sobre procesos cognitivos de carácter subpersonal, a los que no tenemos acceso. Lo que es más, decir que la consecución del estado de cosas deseado es el resultado de nuestro razonamiento no es lo mismo que sostener que todos y cada uno de los procesos que nos permitieron pasar de A a B descansen, por igual, en procesos de deliberación consciente (aunque tampoco cancela la posibilidad de que esto sea así); en realidad, muchos de ellos podrían depender de la puesta en marcha de pautas de acción automáticas.

Es importante hacer notar que en estos contextos de acción tanto la delimitación del problema, como el *procedimiento* de solución y los criterios de evaluación del resultado no son algo que el sujeto construya en el momento mismo en que se enfrenta a la situación, sino que han de ser en-

tendidos como *estándares predefinidos* que sólo hay que poner en práctica. Sin descartar la posibilidad, genuinamente interesante, de que *como individuos* seamos capaces de desarrollar esquemas de acción que estén por completos desligados de nuestro comportamiento actual y sean sólo aplicables en un contextos de actividad posterior —por ejemplo, siguiendo a Popper (1957/1983), se podría argumentar que el diseño de un “experimento crucial” que refute una teoría, o en general de cualquier arreglo experimental, debería estar soportado en algún proceso de este tipo; aunque la idea de esto pueda tener lugar en el terreno meramente individual me parece por demás cuestionable—, estos *sistemas de razonamiento* (o *estilos de razonamiento*, en la terminología empleada por Hackin y Morgan revisada en el capítulo 4) serían de naturaleza social por excelencia: los sistemas lógicos, las matemáticas, la notación musical, las convenciones culinarias, etc.

Ahora bien, es importante aclarar en qué sentido estoy entendiendo que nuestras *respuestas* pueden ser el resultado de nuestro razonamiento. Decimos que una de nuestras acciones *queda explicada* por nuestras razones cuando de la sola consideración del razonamiento se puede inferir la naturaleza de la respuesta o la respuesta misma (es decir, existe una conexión legaliforme entre respuesta y razones), el actor *efectivamente* siguió el razonamiento descrito y, si éste fuera *replicado*, cualquier persona competente en el uso del sistema podría llegar a un resultado estructuralmente equivalente.

Esta caracterización nos lleva a la casi trivial y muy restrictiva idea de que nuestras razones son la causa de *los resultados* de nuestros razona-

mientos. Es decir, nuestras razones dan cuentas de *porqué* “ $E=mc^2$ ”, *porqué* “*cogito ergo sum*” o *porqué* “un pastel de chocolate”, pero no nos dicen nada de porqué Einstein decidió ser físico, porqué Descartes consideraba importante que nos decidiéramos de nuestros prejuicios o porqué el chef no prefirió hacer una gelatina (o, dicho de otro modo, estas explicaciones no pueden ser consideradas en términos motivacionales).

No estoy diciendo aquí que estos razonamientos estén libres de error pero, dado que los criterios que los posibilitan están claramente definidos, incluso estos errores serían reproducibles. Finalmente, se podría argumentar que un razonamiento realmente no es la *causa* de su resultado, sino sólo una especificación de *cómo* (que no porqué) es posible transitar del planteamiento de un problema a su solución. Pienso que esta observación es parcialmente correcta, aunque pasa por alto el hecho de que es intuitivamente evidente que los resultados de nuestros razonamientos son causados, *en algún sentido interesante*, por el hecho de que *efectivamente* hayamos puesto en práctica tales razonamientos. De este modo, se puede decir que, en virtud de que en realidad no soy capaz de reconstruir el razonamiento que llevó a Einstein a postular “ $E = mc^2$ ”, *yo no sé porqué* “ $E = mc^2$ ”.

De cualquier otra acción cuya explicación no pueda ser agotada a partir de la sola consideración del proceso de razonamiento (o de nuestras consciencia reflexiva de la situación) no se puede decir que sea causada por nuestras razones. Esto no equivale a sostener, como parece lo están entendiendo Dijksterhuis, Chartrand y Aarst (2007), que nuestro pensamiento consciente no tenga papel alguno en la determinación del resto de nuestras

conductas, sino sólo que no existe un vínculo causal directo entre nuestra consciencia reflexiva de la situación y nuestro comportamiento. De forma más concreta, la evidencia disponible apunta a que la mayor parte de nuestras acciones en realidad son resultado de la interacción dinámica entre nuestra capacidad de “reinterpretar sistemáticamente las relaciones perceptuales de primer orden en estructuras de orden superior, gobernadas por reglas” (Penn, Holyoak y Povinelli, 2008, p. 111) y nuestros sistemas de percepción-acción (Bermúdez, 2003; Gibson, 1986).

Como lo comentan Baumeister, Masicampo y Vohs (2011, p. 331): “la evidencia de causación consciente de la conducta es profunda, extensa, adaptativa, multifacética y empíricamente sólida. Sin embargo, la causación consciente es frecuentemente indirecta y demorada y depende de la interacción con procesos inconscientes”. A diferencia de Baumeister, Masicampo y Vohs (2011) no considero adecuado hablar de “causación consciente” de la conducta más allá de los (muy escasos) casos que cumplan con la caracterización que acabo de desarrollar, dado que, en estricto sentido de la palabra, ésta sólo podría ser entendida como consecuencia *de la interacción* entre diferentes sistemas cognitivos y no a partir la contribución aislada de *ninguno* de ellos (así, por ejemplo, nos resultaría extraño que alguien afirmara que el sol “causo” la lluvia, aun cuando, efectivamente, el sol juegue un papel muy importante para que ésta se pueda dar; la lluvia y el comportamiento son fenómenos complejos que sólo pueden tener lugar a partir de coordinación de diferentes factores y es tarea de los científicos determinar cómo es que tienen lugar estas interacciones). Con todo, me

parece innegable que nuestra percepción consciente entra en juego en gran parte de nuestros contextos de actividad.

Resulta conveniente distinguir esta propuesta de otras aproximaciones al estudio de la relación entre razones y conducta. A diferencia de lo postulado por las denominadas “teorías duales” (Evans, 2003; 2008), me parece inconcebible la idea de que un sistema cognitivo pueda operar al margen del resto de funcionamiento mental. Esto disuelve el problema de la “resolución del conflicto” entre mecanismos, dado que nuestro propio razonamiento cotidiano (como el resto de nuestras acciones) sólo puede ser entendido como el resultado del funcionamiento coordinado entre sistemas (Damasio, 1994/2006). En la medida en que la determinación del comportamiento sólo puede ser vista como un resultado del funcionamiento global del sistema, la posibilidad de que aparezca un conflicto entre cursos de acción carece de sentido.

Por otro lado, aun aceptando con Penn, Holyoak y Povinelli (2008) que somos capaces de representarnos el mundo como “gobernado por reglas” y de que esta capacidad juega un papel importante en un amplio espectro de nuestra actividad, existe una diferencia muy profunda entre considerar que esta facultad es *el* rasgo definitorio de la cognición humana como un todo y sostener —como yo lo hago, siguiendo a autores como Andy Clark (1998), José Luis Bermúdez (2003) o Michelle Tomasello (1999; 2008)— que ésta es sólo una más entre los diferentes sistemas cognitivos que sustentan nuestro comportamiento. La diferencia más evidente tiene que ver con la forma de explicar la continuidad existente entre nuestra cognición y la de

otros animales: mientras Penn, Holyoak y Povinelli (2008) sostienen que nuestras acciones están mediadas por un sistema representacional unitario esencialmente distinto al presente en otras especies, yo he argumentado que nuestra organización cognitiva sigue el mismo esquema básico presente en otros animales —construido a partir de una serie de mecanismos de percepción-acción—, *apuntalado* (pero no sustituido) por un sistema que nos permite representarnos de forma explícita las relaciones entre eventos. Pretender que, por el sólo hecho de ser los únicos animales que piensan sobre sus propios pensamientos, el estudio científico de nuestro comportamiento demanda que desarrollemos un esquema explicativo radicalmente distinto a aquel del cual echamos mano para dar cuenta de la conducta de otras especies, es como suponer que deberíamos hacer una “psicología del murciélago” sólo por lo peculiar que resulta su sistema de ecolocalización.

Ahora bien, hay que tener mucho cuidado al interpretar la afirmación de que un determinado sistema cognitivo está implicado en una amplia gama de contextos de actividad. En lo específico, no se tiene que confundir esto con la idea de que la integridad de la conducta dependa de este sistema en cada uno de estos escenarios. Por ejemplo, es ampliamente reconocido que somos una “especie visual”, dada la importancia de esta modalidad sensorial para nuestra interacción con nuestro entorno. Sin embargo, nadie estaría dispuesto a afirmar que nuestro comportamiento penda de la capacidad de ver ni que esta facultad esté por igual implicada en todas nuestras acciones. Así, es bien sabido que el sólo hecho de observar un platillo desencadena una serie de reacciones fisiológicas que nos preparan para su in-

gesta. Nuestra experiencia visual es un elemento central de nuestra conducta alimenticia; no obstante, si mañana despertáramos y nos diéramos cuenta que hemos dejado de ver, nuestra capacidad de comer estaría relativamente conservada, mientras que otras áreas de nuestra actividad (como conducir al supermercado a conseguir comida) se verían profundamente afectadas.

Así pues, como Baumeister, Masicampo y Vohs (2011) lo hacen ver: “la investigación futura deberá enfocarse en cómo es que los procesos conscientes e inconscientes interactúan y se complementan los unos a los otros más que en intentar etiquetar cada resultado conductual como dependiente de procesos conscientes o inconscientes” (p. 354). Con respecto a estas pautas de interacción, estos autores comentan:

La evidencia revisada apunta a cuatro grandes conclusiones sobre cómo el pensamiento consciente influye en la conducta. Primero, integra el comportamiento a través del tiempo. [...] Segundo, el pensamiento consciente permite que la conducta de un individuo sea influenciada por factores sociales y culturales. [...] Tercero, el pensamiento consciente es de gran importancia en situaciones que presentan múltiples posibilidades alternativas. [...] Cuarto, la mayoría de la conducta humana, si no es que toda, emerge de una combinación de procesos conscientes e inconscientes. [...] Los esfuerzos pasados por determinar si una conducta dada fue producida por nuestro pensamiento consciente o por nuestro inconsciente pueden estar basados en una falsa dicotomía (pp. 353-354).

Yo no me voy a detener aquí a revisar con detalle todas estas cuestiones; recomiendo ampliamente una lectura minuciosa del trabajo de Baumeister,

Masicampo y Vohs (2011) para los interesados en tener una panorámica clara de la investigación actual en esta área. Mucha es aún la investigación necesaria para esclarecer las pautas de interacción entre razonamiento y conducta. Por el momento, me gustaría concentrarme en un aspecto muy vinculado con el segundo punto que destacan estos autores —la idea de que “el pensamiento consciente permite que la conducta de un individuo sea influenciada por factores sociales y culturales”, p. 353— que, considero, puede establecerse de manera relativamente sólida y tiene profundas implicaciones para la forma en que hasta ahora hemos comprendido la relación entre lenguaje, cognición y comportamiento.

El argumento básico es que, más que ejercer un control directo sobre nuestra conducta, el conocimiento proposicional (*i.e.*, explícito) nos permite modificar nuestros marcos perceptivos, crear nuevos *affordances* que nos permitan proceder a partir de mecanismos de percepción-acción de forma análoga a como lo hace muchas especies, pero ante de contextos de actividad que sólo pueden ser entendidos como *construcciones culturales*. Dicho de otro modo, pensar en la estructura del mundo nos permite modificar la forma en que percibimos el mundo para así ¡poder actuar en él sin tener que pensar tanto cuando actuamos!

La idea de que somos capaces de cambiar (o incluso crear) nuestras propios *affordances* debe de ser evaluada en su justa dimensión. Realmente, cualquier organismo capaz de aprender es, de hecho, capaz de modificar su percepción de la realidad. No obstante, sostengo, nuestra especie posee un sistema cognitivo —nuestra capacidad metarrepresentacional, ésta sí,

única en el reino animal— que nos permite, a partir de su interacción con el resto de nuestra arquitectura mental, llevar esta potencialidad a niveles inobservados en el mundo natural antes de la aparición de los humanos anatómicamente modernos, hace alrededor de 100,000 años (Mithen, 1998). Consideremos esta idea con detalle.

Somos una especie social, no sabe duda, y diversos autores (por ejemplo, Bruner, 1990, Tomasello, 1999; Vygotsky, 1934/1986) han argumentado que para comprender aquello que diferencia a nuestro pensamiento del de “otras especies” hemos de determinar cómo es que nuestra vida social se ve posibilitada por nuestras capacidades cognitivas. Con todo, decir que animales sociales no es mucho; lo son por igual las abejas, las hormigas y los chimpancés. Debe de haber algo en nuestra cognición social que pueda ser visto como la esencia misma de nuestra condición como humanos.

En el capítulo 3 revise la propuesta de autores como Alison Gopnik y Andrew Meltzoff (1999; Gopnik y Wellman, 1994) o Susan Carey (2009; Carey y Johnson, 2000) quienes sostienen que el conocimiento cotidiano se organiza a partir *teorías implícitas* (es decir, como sistemas de creencias que nos permiten organizar e inferir los eventos del mundo a partir de la información de la que previamente disponemos). El conjunto de esta área de investigación, englobada bajo la etiqueta genérica de la “teoría de la teoría”, se vio muy impulsada por el trabajo pionero de Premack & Woodruff (1978, véase también Premack y Premack, 1988) sobre la capacidad de los chimpancés para comprender estados mentales. Aunque es importante aclarar que Premack & Woodruff (1978) no defienden la idea de que los

chimpancés puedan representarse el mundo por medio de sistemas teóricos plenamente desarrollados, sino sólo que pueden entender la conducta de sus cuidadores humanos como guiada por sus estados mentales.

Si bien la repuesta de Premack & Woodruff (1978) fue que, en efecto, sus chimpancés podían comprender que la conducta de una persona estaba gobernada por sus intenciones y creencias, muchas investigaciones posteriores cuestionaron esta conclusión y ofrecieron explicaciones alternativas para sus resultados (véase por ejemplo, Call y Tomasello, 2008; Penn y Povinelli, 2007a). De estas discusiones se desprende la idea de que, aunque en efecto muchos animales son muy sensibles a las acciones de otros organismos (en especial a las de sus congéneres), sólo los humanos somos capaces de comprender el comportamiento de los otros en términos de *actitudes proposicionales* —i.e., “como [un tipo de] estructuras de orden superior, gobernadas por reglas” (Penn, Holyoak y Povinelli, 2008, p. 111). Es muy revelador que la investigación en cognición comparada llegara a apuntalar la misma noción de “psicología popular” prevaleciente en filosofía e inspirada en el trabajo de Russell (1912/1997). Bermúdez (2003a) describe esta posición en los siguientes términos:

Casi cualquiera concede que nos ajustamos a la conducta de nuestros compañeros seres humanos atribuyéndoles creencias y deseos (u otras actitudes proposicionales), asumiendo que ellos actuaran en una forma que es más o menos racional dados esos deseos y creencias, para entonces resolver el curso de acción que esta asunción de racionalidad parece dictar. Por supuesto, cualquiera estará de acuerdo que no hacemos esto de manera *consciente*, pero es usual pensar que, en la gran mayoría de

los casos, esta práctica de atribución, explicación y predicción procede tácitamente (p. 26).

De este modo, todas las teorías de la comunicación inspiradas en la tradición griceana (Grice, 1989; Sperber y Wilson, 1995) parten del supuesto de que el significado lingüístico y el entendimiento entre personas dependen de la capacidad de comprender las *intenciones conversacionales* del interlocutor. Los defensores de la “teoría de la teoría” extienden esta noción a la totalidad de las interacciones entre humanos, defendiendo que éstas sólo son posibles gracias a nuestra capacidad de formarnos una *teoría de la mente* de los demás. (“Para los defensores de la teoría de la teoría, el entendimiento social implica esencialmente un razonamiento subsuntivo. La psicología popular es una cuestión de subsumir patrones de conducta bajo generalizaciones y deducir las consecuencias relevantes” Bermúdez, 2003a, p. 44).

Aunque algunos autores como Paul L. Harris (2000/2005) han argumentado que postular la existencia de un sistema teórico que guie nuestras interacciones sociales es, cuando menos, exagerado y que la adscripción de actitudes proposicionales muy bien podría operar por medio de un sistema cognitivo que nos permitiera *simular* en nuestra mente aquello que nosotros haríamos si nos encontráramos en la posición de la persona con la que interactuamos (básicamente, una cuestión de “ponerse en los de los otros”), es importante notar que tanto la “teoría de la teoría” como la “teoría de la simulación” aceptan que los intercambios sociales en esencia operan por

medio de la atribución de creencias y deseos. Lo que separa a ambas posiciones es sólo: (1) cómo cada una entiende que tiene lugar la atribución de actitudes proposicionales, y (2) la forma en que se concibe el paso de esta atribución a las explicaciones/predicciones del comportamiento (Bermúdez, 2003a).

¿Hemos pues arribado al criterio que nos permita comprender lo que hay de peculiar en nuestra cognición? ¿Es nuestra capacidad de adscribir actitudes proposicionales la esencia del pensamiento humano? Pienso que estos cuestionamientos deben ser respondidos en dos diferentes niveles. Por una parte, efectivamente hay buenas razones para pensar que somos los únicos animales capaces de comprender a otros organismos como *agentes intencionales* (Penn y Povinelli, 2007a). Sin embargo, también puede establecerse de forma relativamente sólida que, aunque podamos formarnos una representación explícita de nuestros contextos de acción, ésta no siempre antecede a nuestra conducta —*i.e.*, no puede estar causalmente vinculada con ella (Bargh, 2005; Dijksterhuis, Chartrand y Aarst, 2007, Wilson, 2002). Pretender que es posible establecer una distinción tajante entre el nivel “social” y el nivel “no-social” de nuestra actividad tampoco se presenta como una opción viable, dado, pienso, esto sólo nos llevaría a una dicotomía artificial más, suponiendo, sin conceder, que de entrada fuera posible etiquetar cada una de nuestras posibles conductas en alguna de estas categorías (¿dónde ubicaríamos, por ejemplo, nuestra capacidad de jugar villar o conducir un fórmula 1?).

No pienso que nuestras interacciones sociales puedan ser simplemente separadas de este esquema general de explicación, sino que, de hecho, pueden resultar como muy buenos ejemplos de él. Pensemos en cualquiera de nuestras conductas cotidianas que implican interactuar con otras personas: abordar el transporte público, comprar un café, ir a un restaurante o comprar un boleto para entrar a un museo. ¿Es necesario que desarrollemos un sofisticado esquema de adscripción de deseos y creencias cuando interactuamos con la persona en la taquilla del metro o el museo? No creo que haya razón alguna para suponer que esto sea así. Como lo comenta Bermúdez (2003a):

Ordenar comida en un restaurante y comprar carne en una carnicería son situaciones tan rutinarias [...] en que el sólo hecho de identificar los roles sociales proporciona una comprensión suficiente de la situación como para permitirnos predecir la conducta de los otros participantes y comprender por qué actúan en la forma en que lo hacen. [...] El punto es que estos son casos donde nuestra comprensión de los individuos y sus conductas son parasitarias de nuestra comprensión de las prácticas sociales en que éstas tienen lugar. Por supuesto que este entendimiento de las prácticas sociales no pasa por un dominio de algún tipo de teoría primitiva. Aprendemos a través de la experiencia que ciertas claves sociales están correlacionadas con ciertos patrones conductuales de parte de los otros y ciertas expectativas de ellos, a su vez, sobre cómo deberíamos de actuar (pp. 43-44).

Lo interesante de este tipo de situaciones —que, habríamos de reconocer, abarcan la mayor parte de nuestras interacciones cotidianas— es que re-

sultan ejemplos particularmente claros en que la percepción de patrones ambientales nos lleva a la emisión de conductas adecuadas sin la necesidad de representarnos explícitamente toda la situación. Es decir, gran parte de nuestra vida social podría muy bien ser explicada en términos puramente perceptuales (Bermúdez, 2003), lo cual hace evidente que el dominio de aplicación de nuestra “psicología popular” debe ser mucho más estrecho de lo que tradicionalmente se piensa en filosofía y ciencia cognitiva. Al respecto Bermúdez (2003a) comenta:

Ex hyphotesi, sin embargo, en entendimiento de los roles sociales y rutinas asociadas con razonamientos basados en el contexto está restringido a situaciones sociales familiares. ¿Qué pasa cuando nos encontramos ante una situación social que no nos es familiar? [...] Es natural afirmar que en este punto ya hemos arribado al dominio propio de la psicología popular. Uno puede pensar que el entendimiento social es una herramienta compleja para negociar el mundo social. Éste es frecuentemente transparente, fácilmente comprensible en términos de marcos de acción (*frames*), roles y rutinas sociales. [...] No necesitamos echar mano de la maquinaria de atribución característica de la psicología popular para navegar a través del mundo social, para ajustarnos a las necesidades y requerimientos de la otra gente y para coordinar actividades de manera exitosa. Pero en ocasiones el mundo social se vuelve opaco. Nos encontramos en interacciones sociales donde no es obvio qué está sucediendo, que no pueden ser asimiladas de manera sencilla a situaciones prototípicas, donde no podemos determinar qué hacer simplemente basándonos en interacciones previas con otros participantes. Y es en este punto en que necesitamos del tipo de pensamiento metarrepresentacional característico de la psicología popular —no como fundamento de nuestro entendimiento social, sino como el

último recurso del que disponemos cuando todos los mecanismos estándar de comprensión social y coordinación interpersonal resultaron insuficientes (pp. 46-47).

Esta intuición queda capturada en la máxima propuesta por Bermúdez (2003) según la cual: “todo pensamiento que involucra acenso intencional (a grandes rasgos, todo pensamiento que involucra pensar sobre pensamientos) requiere la capacidad para acenso semántico (a grandes rasgos, la capacidad de pensar sobre palabras)” (p. 151). De esta forma, se aclara el infranqueable papel de del lenguaje en nuestra cognición: sólo a partir de éste podemos representarnos el mundo mental, tanto de los otros como propio.

Resumiendo, la idea de que nuestra psicología popular —y, con ella, toda nuestra consciencia reflexiva de la realidad— es, en esencial, un sistema cognitivo implicado en la *negociación social* de diferentes aspectos del mundo resulta —al menos así me lo parece— totalmente consistente con la evidencia empírica de que disponemos hasta este punto. Opiniones similares han sido expuestas de manera reciente por Baumeister y Masicampo (2010) y Mercier y Sperber (2011).

Todas estas propuestas coinciden en la necesidad de invertir la forma en que tradicionalmente se ha entendido la relación entre percepción y razonamiento: mientras que las propuestas teóricas imperantes hasta este momento defienden que nuestros sistemas perceptuales nos proporcionan

la información necesaria para razonar sobre la situación y actuar de forma adecuada ante ella, el panorama conceptual que he reconstruido a lo largo de la segunda sección de esta tesis apunta a que, en realidad, nuestra consciencia reflexiva del mundo muchas veces no tiene un papel causal directo sobre nuestras acciones, sino que nos permite cambiar la forma en que nos representamos la realidad, crear nuestras propias *affordances* y aprovechar los mismos mecanismos de percepción-acción que tan eficientes se han mostrado en la regulación conductual de los organismos a lo largo de la evolución biológica.

Pienso que esta aproximación tiene la bondad de dar cuenta de la gran continuidad existente entre nuestra cognición y la de otros animales al unificar la explicación del comportamiento bajo un esquema general que entiende que las acciones de un organismo sólo pueden ser entendidas a la luz de sus capacidades perceptivas y que ve a la percepción como un fenómeno complejo de integración dinámica entre diferentes sistemas cognitivos.

Por otra parte, considero que una propuesta de esta naturaleza también sienta las bases para una aproximación genuinamente naturalista al estudio del conocimiento humano, al permitirnos entenderlo ya no sólo como la acumulación de saberes proposicionales, sino como la construcción conjunta de nuevos marcos perceptuales. Una aproximación de este tipo ha sido recientemente puesta en práctica por Ángeles Eraña y Sergio Martínez (2004) para abordar el desarrollo del conocimiento científico. Como comentan estos autores:

[El] despliegue de las habilidades prácticas implicadas en la cognición [tanto en la ciencia como en la vida diaria] no consiste simplemente en un procesamiento explícito de símbolos; es más bien el resultado de una serie de acciones coordinadas que tienen lugar en el ambiente normativo de un complejo de situaciones, donde las normas frecuentemente no están explícitas (sino implícitas) en las prácticas subyacentes (p. 725).

Por desgracia, una revisión detallada de las implicaciones de esta propuesta es algo que ya está más allá de los alcances de este trabajo.

CONCLUSIONES

EL CONOCIMIENTO COMO FENÓMENO NATURAL

Las conclusiones de este trabajo se pueden resumir en los siguientes términos:

1. La relación entre psicología y epistemología siempre ha sido profunda y multifacética. El explosivo desarrollo teórico que la psicología experimentó a lo largo del siglo XX es simplemente incomprensible sin hacer mención de la filosofía positivista de la ciencia y de sus concepciones del lenguaje y el conocimiento. Ignorar estas interacciones sólo nos puede llevar a la confusión y perpetuación de viejos errores, mientras que hacerlas explícitas nos permitiría construir un contexto de discusión fructífero donde los profesionales de diferentes áreas puedan abonar con sus habilidades y conocimientos a nuestra comprensión de los fenómenos cognitivos.

2. La idea de que la mente y la ciencia pueden ser descritas a partir de las mismas categorías conceptuales de las que uno echa mano a la hora de caracterizar los sistemas lingüísticos está en el núcleo mismo de la filosofía analítica, sin embargo, no se restringe a ella. Ideas similares pueden ser identificadas como los fundamentos de las aproximaciones narrativas a estudio de los fenómenos humanos. Lo que cambia entre una y otra perspectiva es la caracterización del lenguaje que se considera más adecuada, pero no el supuesto de que tanto la mente como la ciencia pueden ser entendidas como sistemas simbólicos.

3. Tanto la teoría computacional de la mente como el análisis experimental de la conducta (AEC) tienen sus fundamentos conceptuales en las discusiones filosóficas sobre la naturaleza del lenguaje y del conocimiento científico que tuvieron lugar durante la primera mitad del siglo XX. La teoría computacional de la mente es incomprensible sin los desarrollos lógicos que permitieron el nacimiento de la inteligencia artificial (IA) así como sin la idea de que los lenguajes naturales pueden ser equiparados a los lenguajes formales. Por otra parte, el AEC puede ser visto como una expresión más de la idea de que las teorías científicas deben ser expresadas por medio de sistemas formales, como la lógica y la matemática. En este sentido, más que posiciones antagónicas, la teoría computacional de la mente y el AEC se complementan y regulan mutuamente.

4. Es importante no confundir las teorías psicológicas con los fenómenos psicológicos que describen. En virtud de que ambos son concebidos en términos lógicos, suele ser muy común que se equipare a las descripciones

formales de fenómenos psicológicos con algoritmos cognitivos efectivamente instaurados en la mente/cerebro. Sin embargo, un análisis detallado hace evidente que mientras los algoritmos cognitivos sólo pueden ser entendidos como reglas de procesamiento que guían la transición entre tipos de representaciones y que están vinculados estrechamente con elementos específicos de la arquitectura neural/cognitiva, las descripciones formales de fenómenos psicológicos son de naturaleza más general y guardan un estatus ontológico totalmente distinto al de los algoritmos, siendo comprensibles sólo a la luz de nuestras prácticas epistemológicas y a partir de su interacción con otras herramientas materiales y conceptuales de las que echamos mano en nuestros intentos por entender la naturaleza.

5. Contrario a lo defendido por autores como Fodor y Pylyshyn (1988), la sola especificación de nuestra arquitectura cognitiva no agota la explicación del carácter sistemático de nuestro pensamiento. Es decir, una descripción de los cálculos que lleva a cabo nuestro cerebro cuando hacemos frente a las demandas del ambiente no es, por sí misma, una teoría de la racionalidad (esto es porque, por definición, una máquina universal podría simular cualquier manipulación de símbolos, por lo que la arquitectura mental de un “organismo” programado para actuar constantemente de forma “irracional” (a la luz de algún criterio preestablecido) seguiría siendo sintácticamente especificable; *i.e.*, computable \neq racional). Con todo, es un hecho que entender a la mente como un sistema que manipula sistemáticamente símbolos constituyó un profundo impulso a la legitimación de la noción cartesiana de racionalidad, que equipara ser racional con pensar a

partir de las reglas de la lógica (y, posteriormente, de la matemática).

5. No obstante, no todas las conductas adaptativas (que podrían ser calificadas como racionales a partir de sus consecuencias para el organismo que las emite) pueden ser entendidas como consecuencias de procesos racionales de pensamiento. De este modo, el gran reto para la ciencia cognitiva está precisamente en determinar la relación existente entre razones y acciones.

6. Autores como José Luis Bermúdez (2003) o Penn, Holyoak y Povinelli (2008) argumentan que es posible dar cuenta del comportamiento de los organismos carentes de lenguaje a partir de su capacidad de percibir sus contextos de actividad como altamente complejos y estructurados, sin apelar a genuinos procesos de toma de decisiones y, en consecuencia, sin apelar a procesos plenamente racionales.

7. Si aceptamos lo anterior, la adscripción de procesos racionales se restringiría a los organismos lingüísticamente competentes, por lo que habría de ser entendida como propiedad característica de la cognición humana.

8. No obstante, es mucha la evidencia que apunta a que nuestras razones no siempre funcionan como causas de nuestras acciones. Primero, muchas veces no somos conscientes de qué estímulos ejercen un control efectivo sobre nuestro comportamiento. Segundo, en ocasiones nuestra conducta precede a nuestros razonamientos.

9. De hecho, gran parte de la conducta humana puede ser explicada a partir de los mismos mecanismos de percepción-acción que, se presume, sustentan la mayoría de la conducta inteligente de los organismos carentes de lenguaje. Esto me permite afirmar que es posible un esquema explicati-

vo general que, partiendo de la idea de que la percepción es un fenómeno complejo orientado al control directo del comportamiento, nos permita dar cuenta, por igual, del grueso de las acciones de todos los organismos vivos.

10. Con lo anterior en mente, ha de ser claro que si el razonamiento juega un papel en nuestra vida mental, éste es mucho más restringido de lo que tradicionalmente ha sido aceptado por las teorías psicológicas clásicas. Siguiendo a Bermúdez (2003; 2003a), Mercier y Sperber (2011) y Baumeister y Masicampo (2010), defiendo la idea de que el razonamiento debe ser visto como una capacidad cognitiva metarrepresentacional que nos permite modificar los esquemas perceptuales que guía nuestro comportamiento a partir de nuestras interacciones comunicativas con otras personas. Dicho de otra forma, nuestra capacidad de pensar sobre pensamientos nos permite *negociar socialmente* nuestra propia realidad, para crear contextos de acción más allá de los que nos fueron biológicamente dados.

A lo largo del (difícil) proceso de elaboración de esta tesis, mis concepciones sobre la cognición y la relación entre el razonamiento y la conducta se vieron radicalmente modificadas. Sin embargo, hubo una idea que siempre tuve muy presente y que podría ser vista como la motivación última para el desarrollo de este trabajo: *estoy convencido de que es necesaria (y posible) una teoría naturalista del conocimiento humano*, una teoría que nos permita dar cuenta por igual del desarrollo de nuestros saberes cotidianos y de sistemas tan ricos y complejos como nuestra ciencia y nuestro arte.

Siempre he considerados erróneos los intentos por explicar el fenómeno del conocimiento a partir de las capacidades cognitivas de los miembros aislados de nuestra especie (aunque también considero que una comprensión cabal de de estas capacidades es un prerequisite *sine qua non* cualquier teoría del conocimiento resultaría trivial). Entiendo que nuestro conocimiento no puede ser entendido más que en términos históricos, no como consecuencia del desarrollo de las mentes individuales, sino sólo a partir de la evolución de las pautas de interacción *entre mentes*.

Así lo indican hechos tan concretos como (1) la observación realizada por Mithen (1998) de que, mientras que la aparición del los humanos anatómicamente modernos se remonta a aproximadamente 100,000 años, hace sólo 40,000 años (la mitad de nuestro tiempo como especie) que empezaron a aparecer las manifestaciones más características de la cultura humana; (2) el que hasta antes de la aparición de los sistemas de escritura (hace sólo aproximadamente 5,000 años) los intercambios de información se restringían a aquello que podría ser compartido cara a cara (Clark, 1998; Olson, 1998), (3) el hecho de que nuestra concepción moderna de ciencia no se remonte más allá de 500 años atrás en nuestra historia (Olson, 1998) o (4) el que hace tan sólo 20 años resultara imposible imaginar las consecuencias que una herramienta como la Internet iba a tener para nuestra vida cotidiana (Carr, 2010).

Aún sin negar que estos eventos hayan impactado la organización de nuestro cerebro, concuerdo con Clark (1997; 1998) en que no hay razones para suponer que estos cambios haya sido radicales: seguimos actuando a

partir del reconocimiento de patrones, aunque esos patrones enmascaren una complejidad social cada vez mayor. En este sentido, la comprensión de nuestra organización cognitiva (a la que prácticamente se restringe esta tesis) es sólo la mitad de camino por recorrer para comprender la naturaleza del conocimiento.

Tener que reconocer que todo lo que he podido decir en este trabajo no puede ser visto más que como una introducción (por demás imprecisa e incompleta) al problema del conocimiento me provoca sentimientos encontrados. Por una parte, estoy insatisfecho de no haber siquiera podido atisbar una solución al problema central que tenía en mente cuando comencé este trabajo. Por otro lado, al tener que concluir, puedo decir que estoy más convencido que nunca de la complejidad y la importancia de concebir al conocimiento como un fenómeno natural y de que es posible, aunque no resultará fácil, aproximarse a él con curiosidad científica.

Bibliografía:

- Abadie, A. y Gay, S. (2006). The impact of presumed consent legislation on cadaveric organ donation: A cross-country study. *Journal of Health Economics*, 25, pp. 599-620.
- Ackerman, J.M., Nocera, C.C. y Bargh, J.A. (2010). Incidental Haptic Sensations Influence Social Judgments and Decisions. *Science*, 328, 1712-1715.
- Armstrong, D.M. (1970/1980). The Nature of Mind. En N. Block (Ed.), *Readings in Philosophy of Psychology*, Vol. 1. (pp. 191-199). Cambridge Mass.: Harvard University Press.
- Bargh, J.A. (2005). Bypassing the Will: Toward Demystifying the Nonconscious Control of Social Behavior. En R.R. Hassin, J.S. Uleman y J.A. Bargh (Eds.), (pp. 37-58).
- Bargh, J.A. (2007) *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes*. New York: Psychology Press.
- Bargh, J.A. y Shalev, I. (2011). The Substitutability of Physical and Social Warmth in Daily Life. *Emotion*, Advance online publication. doi: 10.1037/a0023527.

- Baumeister, R.F. y Masicampo, E.J. (2010) Conscious Thought Is for Facilitating Social and Cultural Interactions: How Mental Simulations Serve the Animal-Culture Interface. *Psychological Review*, 117(3), 945-971.
- Baumeister, R.F., Masicampo, E.J. y Vohs, K.D. (2011). Do Conscious Thoughts Cause Behavior? *Annual Review of Psychology*, 62, 331-361.
- Bennett, M., y Hacker, P. (2003). *Philosophical Foundations of Neuroscience*. Oxford: Blackwell.
- Bermúdez, J.L. (2003). *Thinking without Words*. New York: Oxford University Press.
- Bermúdez, J.L. (2003a). The Domain of Folk Psychology. En A. O'Hear (Ed.), *Minds and Persons* (pp. 25-48). Cambridge: Cambridge University Press.
- Brembs, B. (2003). Operant conditioning in invertebrates. *Current Opinion in Neurobiology*, 13, 710-717.
- Bruner, J. (1990). *Actos de significado: Más allá de la revolución cognitiva*. Madrid: Alianza.
- Brun, G. y Kuenzle, D. (2008). Introduction: A New Role for Emotions in Epistemology? En. G. Brun, U. Doğuoğlu y D. Kuenzle (Eds.), *Epistemology and Emotions* (pp. 1-31). Hampshire: Ashgate
- Buss, D.M. (2008). *Evolutionary psychology: The new science of the mind* (3° Ed.). Boston. Pearson Education.
- Call, J. y Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, 12(5), 187-192.
- Carey, S. (2009). *The Origin of Concepts*. New York: Oxford University Press.
- Carey, S. y Johnson, S.C. (2000). Metarepresentation and Conceptual Change: Evidence from Williams Syndrome. En D. Sperber (Ed.), *Metarepresentations: A Multidisciplinary Perspective* (pp. 225-264). New York: Oxford University Press.
- Carnap, R. (1932/2002). Psychology in Physical Language (Excerpt). En D.J. Chalmers (Comp.), *Philosophy of mind: Classical and Contemporary Readings* (pp. 39-44). New York: Oxford University Press.
- Carnap, R. (1937). *The Logical Syntax of Language*. Londres: Routledge.

- Carr, N. (2010). *The Shallows: What the Internet Is Doing to Our Brains*. New York: W.W. Norton & Company.
- Cartwright, N. (1980/1991). The Reality of Causes in a World of Instrumental Laws. En R. Boyd, P. Gasper y J.D Trout (Eds.), *The Philosophy of Science* (pp. 178-386). Cambridge, Mass: The MIT Press.
- Chomsky, N. (1957/2002). *Syntactic Structures (2° ed)*. New York: Mouton dc Gruyter.
- Chomsky, N. (1959). A Review of B. F. Skinner's *Verbal Behavior*. *Language*, 35(1), 26-58.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, Mass: The MIT Press.
- Chomsky, N. (2009). *Cartesian Linguistics: A Chapter in the History of Rationalist Thought (3° Ed)*. New York: Cambridge University Press.
- Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, Mass: The MIT Press.
- Churchland, P.M. (1981) Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy*, 78(2), 67-90.
- Cosmides, L y Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. En L.A. Hirschfeld y S.A Gelman (Eds.), pp. 85-116.
- Cosmides, L. y Tooby, J. (1997). *Evolutionary Psychology: A primer*. Recuperado el 02/10/2009 de <http://www.psych.ucsb.edu/research/cep/primer.html>
- Damasio, A. (1994/2006). *L'erreur de Descartes: La raison des émotions*. Paris: Odile Jacob.
- Davidson, D (1963/2001). Actions, reasons and causes. En *Essays on Actions and Events* (pp. 3-19). Oxford: Clarendon Press.
- de Saussure, F. (1916/2008). *Curso de lingüística general*. México: Fontamara.
- de Waal, F. (2006) *Primates and Philosophers: How Morality Evolved*. Princeton: Princeton University Press.
- Descartes, R. (1637/2001). *Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences*. Mozambook.

- Dijksterhuis, A., Chartrand, T.L. y Aarst, H. (2007). Effects of Priming and Perception on Social Behavior and Goal Pursuit. En J.A. Bargh (Ed.), (p. 51-131).
- Dijksterhuis, A. y van Olden, Z. (2006). On the benefits of thinking unconsciously: Unconscious thought can increase post-choice satisfaction. *Journal of Experimental Social Psychology*, 42, 627–631.
- Dilthey, W. (1890/1980). *Introducción a las Ciencias del Espíritu*. Madrid: Alianza Editorial.
- Domingo, J.M. (2003). El proyecto modular de Jerry Fodor (o sobre el porvenir de otra ilusión). *Anuario de Psicología*, 34(4), 364-371.
- Dutton, D.G. y Aron, A.P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology*, 30(4), pp. 510-517.
- Eraña, A. (2006). Las teorías intuitivas ¿son módulos. *Revista de Filosofía de la Universidad de Costa Rica*, 44(111), 85-99.
- Eraña, A. y Martínez, S.F. (2004). The Heuristic Structure of Scientific Knowledge. *Journal of Cognition and Culture*, 4(3), 701-729.
- Evans, J.St.B.T (2003). In two minds: dual-process accounts of reasoning. *TRENDS in Cognitive Sciences*, 7(10), 454-459.
- Evans, J.St.B.T (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, 59, 255-278.
- Fodor, J. (1975). *The Language of Thought*. New York: Tomas y Crowell Company.
- Fodor, J. (1983). *The Modularity of Mind*. Cambridge, Mass.: The MIT Press.
- Fodor, J. y Pylyshyn, Z. (1988) Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-71.
- Gardner, H. (1985). *The Minds New Science: A History of the Cognitive Revolution*. New York. Basic Books.
- Gibson, J.J. (1986). *The Ecological Approach to Visual Perception*. New York: Psychology Press.
- Gigerenzer, G. (2004). Fast and Frugal Heuristics: The Tools of Bounded Rational. En D.J. Koehler y N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making* (pp. 62-88). Malden: Blackwell.

- Gigerenzer, G. (2007). *Gut Feelings: The Intelligence of the Unconscious*. Londre: Viking Penguin.
- Gigerenzer, G. y Brighton, H. (2009). Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science*, 1, 107–143.
- Gigerenzer, G. y Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review Psychology*, 62, 451–482.
- Goldberg, E. (2002). *El cerebro ejecutivo: lóbulos frontales y mente civilizada*. Barcelona: Crítica.
- Gopnik, A. y Meltzoff, A.N. (1999). *Palabras, pensamientos y teorías*. Madrid: Visor.
- Gopnik, A. y Tenenbaum, J.B. (2007). Bayesian networks, Bayesian learning and cognitive development. *Developmental Science*, 10(3), 281-287.
- Gopnik, A. y Wellman, H.M. (1994). The theory theory. En L.A. Hirschfeld y S.A Gelman (Eds.), pp. 257-293.
- Grice, P. (1989). *Studies in the Way Of Words*. Cambridge, Mass.: Harvard University Press.
- Griffin, D.R. (1994). *Animal minds*. Chicago: The University of Chicago Press.
- Harris, P.L. (1994). Thinking by children and scientists: False analogies and neglected similarities. En L.A. Hirschfeld y S.A Gelman (Eds.), pp. 294-315.
- Harris, P.L. (2000/2005). *El funcionamiento de la imaginación*. Buenos Aires: Fondo de Cultura Económica.
- Hassin, R.R., Uleman, J.S. y Bargh, J.A. (2005). *The New Unconscious*. New York: Oxford University Press.
- Hauser, M.D. (2006). *Moral Minds: The nature of right and wrong*. Ecco Press.
- Hempel, C.G. (1931/1980). The logical Analysis of Psychology. En N. Block (Ed.), *Readings in Philosophy of Psychology*, Vol. 1. (pp. 14-22). Cambridge Mass.: Harvard University Press.
- Hirschfeld, L.A. y Gelman, S.A. (1994). *Mapping the Mind: Domain Specificity in Cognition and Culture*. New York: Cambridge University Press.

- Hirstein, W. (2005). *Brain Fiction: Self-Deception and the Riddle of Confabulation*. Cambridge, Mass.: The MIT Press.
- Hurley, S. y Nudds, M. (2006) *Rational Animals?* New York: Oxford University Press.
- Johnson, E.J. y Goldstein, D. (2003). Do Defaults Save Lives? *Science*, 332, 1338-13339.
- Jozefowicz, J y Staddon, J.E.R. (2008). Operant behavior. En R. Menzel y J.H. Byrne (Eds.), *Learning and memory, a comprehensive reference, Volume 1: Learning Theory and Behavior* (pp. 75-102). Oxford: Elsevier.
- Kahneman, D. y Tversky, A. (1984). Choices, Values, and Frames. *American Psychologist*, 39(4), 341-350.
- Kandel, E.R. (2008). *In search of memory: The Emergence of a New Science of Mind*. New York: W.W. Norton.
- Keil, F.C. (1994). The birth and nurturance of concepts by domains: The origins of concepts of living things. En L.A. Hirschfeld y S.A Gelman (Eds.), pp. 234-254.
- Knobe, J. y Nichols, S. (2008). *Experimental Philosophy*. New York: Oxford University Press.
- Kuhn, T.S. (1970). *The Structure of Scientific Revolutions* (2° ed.). Chicago: The University of Chicago Press.
- Leslie, A. (1994). ToMM, ToBY, and Agency: Core architecture and domain specificity. En L.A. Hirschfeld y S.A Gelman (Eds.), pp. 119-148.
- Lévi-Strauss, C. (1973/1979). *Antropología Estructural*. México: Siglo XXI.
- Lewis, D.K. (1966/1984). *Un argumento en favor de la Teoría de la Identidad*. México: Instituto de Investigaciones Filosóficas/UNAM.
- Luce, D. (1995). Four Tensions concerning Mathematical Modeling in Psychology. *Annual Review of Psychology*, 46, 1-26.
- Machery, E., Mallon, R., Nichols, S. y Stich, S., (2004/2008). Semantics Cross-Cultural Style. En J. Knobe y S. Nichols (2008, Eds.), pp. 47-71.
- Marr, D. (1982/2010). *Vision: A computational investigation into the human representation and processing of visual information*. Cambridge, Mass.: The MIT Press.

- Mercier, H. y Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34, 57-111.
- Mitchell, C.J., De Houwer, J. y Lovibond, P.F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, 32, 183-246.
- Mithen, S. (1998). *Arqueología de la mente: Orígenes del arte, de la religión y de la ciencia*. Barcelona: Crítica.
- Moravcsik, J.M. (2008). Frege and Chomsky on Thought and Language. *Midwest Studies In Philosophy*, 6(1), 105-124.
- Morgan, M.S. (2009). Modelling as a Method of Enquiry. En M.S. Morgan (en prensa) *The World in the Model: How Economist Work and Think*. Cambridge University Press.
- Moulines, U. (2011). *El desarrollo moderno de la filosofía de la ciencia*. México: UNAM/Instituto de Investigaciones Filosóficas.
- Myung, I.J., y Pitt, M.A. (2001). Mathematical modeling. En J. Wixted (Ed.), *Stevens' handbook of experimental psychology. Vol. 4: Methodology* (pp. 429-459). New York: Wiley.
- Neurath, O. (1913/1983). The lost wanderers of Descartes and the auxiliary motive. En *Philosophical Papers 1913-1946*. Dordrecht: Reidel Publishing Company.
- Neurath, O. (1932/1983). Protocol Statements. En *Philosophical Papers 1913-1946*. Dordrecht: Reidel Publishing Company.
- Neurath, O. (1937). Unified Science and Its Encyclopaedia. *Philosophy of Science*, 4(2), 265-277.
- Olivé, L. (1996). *Razón y Sociedad*. México: Fontamartz.
- Olson, D. (1998). *El mundo sobre el papel: El impacto de la lectura y la escritura sobre la estructura del conocimiento*. Barcelona: Gedisa.
- Penn, D.C., Holyoak, K.J. y Povinelli, D.J. (2008) Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, 31, 109-178.
- Penn, D.C. y Povinelli, D.J. (2007). Causal Cognition in Human and Nonhuman Animals: A Comparative, Critical Review. *Annual Review of Psychology*, 58, 97-118.

- Penn, D.C. y Povinelli, D.J. (2007a). On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. *Phil Trans R Soc B*, 362, 731-744.
- Piaget, J. (1928/2002). *Judgement and Reasoning in the Child*. Routledge.
- Pinker, S. (1997). *How the Mind Works*. New York: Penguin Group.
- Prinz, J. (2008). Empirical Philosophy and Experimental Philosophy. En J. Knobe y S. Nichols (2008, Eds.), pp. 189-208.
- Popper, K.R. (1957/1983). La ciencia: conjeturas y refutaciones. En K.R. Popper (1972/1983), *Conjeturas y refutaciones* (pp. 57-93). Barcelona: Paidós.
- Pozo, J.I. (2003). *Adquisición del conocimiento: Cuando la carne se hace verbo*. Madrid: Morata.
- Premack, D., y Premack, A.J. (1988). *La mente del simio*. Madrid: Debate.
- Premack, D. y Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 4, 515-526.
- Putnam, H. (1963/1983). *Cerebro y conducta*. México: Instituto de Investigaciones Filosóficas/UNAM.
- Putnam, H. (1967/1981). *La naturaleza de los estados mentales*. México: Instituto de Investigaciones Filosóficas/UNAM.
- Putnam, H. (1990). *Realism with a Human Face*. Cambridge, Mass.: Harvard University Press.
- Pylyshyn, Z. (1989/2007). La Computación en Ciencia Cognitiva. En L.A. Pérez-Miranda (Ed.), *Lecturas filosóficas de ciencia cognitiva* (pp. 23-73). Bilbao: Universidad del País Vasco.
- Rachlin, H. (1995). Behavioral economics without anomalies. *Journal of the Experimental Analysis of Behavior*, 64, 397-404.
- Ramachandran, V.S. y Blakeslee, S. (1998). *Phantoms in the brain: probing the mysteries of the human mind*. New York: William Morrow and Company.
- Rescorla, R.A. y Wagner, A.R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. En A.H. Black y W.F. Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory* (pp. 64-99). New York: Appleton-Century-Crofts.

- Ricoeur, P. (1970/1986). Qu'est-ce qu'un texte? Expliquer et comprendre. En P. Ricoeur (1986), pp. 137-182.
- Ricoeur, P. (1971/1986). Le modele du texte : l'action sensée considérée comme un texte. En P. Ricoeur (1986), pp. 183-212.
- Ricoeur, P. (1977/1986). Expliquer et comprendre. Sur quel~ues connexions remarquables entre la theorie du texte, la théorie de l'action et la théorie de l'histoire. En P. Ricoeur (1986), pp. 161-182.
- Ricoeur, P. (1986). *Du texte à l'action. Essais d'herméneutique II*. Paris: Éditions du Seuil.
- Robinson, E.S. (1939). The Unity of Science Movement and the Social Sciences. *Proceedings of the Oklahoma Academy of Science*, 20, 129-130.
- Russell, B. (1912/1997). *The Problems of Philosophy*. Londres: Oxford University Press.
- Ryle, G. (1949/2005). *El concepto de lo mental*. Barcelona: Paidós.
- Sacks, O. (1988). *L'homme qui prenait sa femme pour un chapeau et autres récits cliniques*. Paris: Éditions du Seuil.
- Savage-Rumbaugh, S., Shanker, S.G. y Taylor, T.J. (1998). *Apes, Language, and the Human Mind*. New York: Oxford University Press.
- Schwartz, B. (2004). Why Mores is Less: The paradox of choice. HarperCollins e-books.
- Shettleworth, S.J. (2010). *Cognition, Evolution and Behavior* (2° ed.). New York: Oxford University Press.
- Skinner, B.F. (1938/1975). *La conducta de los organismos: Un análisis experimental*. Barcelona: Fontanella.
- Skinner, B.F. (1957). *Verbal Behavior*. New York: Appleton-Century-Crofts.
- Smart, J.J.C. (1959). Sensations and brain processes. *The Philosophical Review*, 68(2), 141-156.
- Sperber, D. (1994). The modularity of thought and the epidemiology of representations. En L.A. Hirschfeld y S.A Gelman (Eds.), pp. 37-67.
- Sperber, D. (1996). *Explaining Culture: A Naturalistic Approach*. Oxford: Blackwell.

- Sperber, D. y Wilson, D. (1995). *Relevance: Communication and Cognition*. Oxford: Blackwell.
- Sutherland, S. (1996). *Irracionalidad: el enemigo interior*. Alianza: Madrid.
- Todd, P.M. y Gigerenzer, G. (2000). Précis of Simple heuristics that make us smart. *Behavioral and Brain Sciences*, 23, 727-780.
- Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, Mass.: Harvard University Press.
- Tomasello, M. (2005). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, Mass.: Harvard University Press.
- Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, Mass.: The MIT Press.
- Turing, A.M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, Series 2*, 43, 544 - 546.
- Turing, A.M (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433-460.
- Tversky, A y Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185, 1124–1130.
- van Fraassen, B. (1977). The Pragmatics of Explanation. *American Philosophical Quarterly*, 14, 143-50.
- Vygotsky, L. (1934/1986). *Thought and Language*. Cambridge, Mass: The MIT Press.
- Wasserman, E.A. y Zentall, T.R. (2009). *Comparative Cognition: Experimental Explorations of Animal Intelligence*. New York: Oxford University Press.
- Wason, P.C. y Shapiro, D. (1971). Natural and Contrived Experience in a reasoning Problem. *Quarterly Journal of Experimental Psychology*, 23, 63-71.
- Watson, J.B. (1930/1970). *Behaviorism*. New York: W.W. Norton.
- Wertsch, J.V. (1988). *Vygotsky and the Social Formation of Mind*. Cambridge, Mass: The MIT Press.
- Weinberg, J.M, Nichols, S. y Stich, S., (2001/2008). Normativity and Epistemic Intuitions. En J. Knobe y S. Nichols (2008, Eds.), pp. 17-45.

- Wilson, T.D., Lisle, D.J., Schooler, J.W., Hodges, S.D., Klaaren, K.J. y LaFleur, S.J. (1993). Introspecting About Reasons Can Reduce Post-Choice Satisfaction. *Personality and Social Psychology Bulletin*, 19(3), 331-339.
- Wilson, T.D. y Schooler, J.W. (1991). Thinking Too Much: Introspection Can Reduce the Quality of Preferences and Decisions. *Journal of Personality and Social Psychology*, 60(2), 181-192.
- Wilson, T.D. (2002). *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, Mass.: Harvard University Press.
- Williams, L.E. y Bargh, J.A. (2008). Experiencing Physical Warmth Promotes Interpersonal Warmth. *Science*, 332, 606-607.
- Wittgenstein, L. (1953/1997). *Philosophical Investigations*. Malden, Mass. Blackwell.