



**Universidad Autónoma Metropolitana  
Unidad Iztapalapa  
Posgrado en Humanidades  
Línea Historia y Filosofía de la Ciencia**

***El evolucionismo de Dennett: las teorías del ego y de la conciencia***

**Tesis de Maestría  
(Idónea comunicación de resultados)**

**Que presenta  
Lic. Víctor Manuel Romero Sánchez**

**Director  
Dr. Jorge Martínez Contreras**

**Jurado  
Dr. Cuauhtémoc Lara Vargas  
Dr. Jonatan García Campos  
Dr. José Luis Vera Cortés**

**Julio de 2010**

## Índice

Introducción	3
Capítulo I La teoría de la conciencia	8
I.1 El Teatro Cartesiano	8
I.2 El modelo del Teatro Cartesiano a prueba	10
I.3 El surgimiento de los modelos de la conciencia en Dennett	20
I.4 Los modelos y las metáforas de la conciencia	25
I.5 La evolución de la conciencia	33
Capítulo II La teoría del ego	39
II.1 Las metáforas del ego	41
II.2 Respecto del ego como narración	46
II.3 La evolución del ego	51
Capítulo III Los experimentos de Libet	59
III.1 Medición y arribo de los contenidos a la “conciencia”	61
III.2 Breves consideraciones sobre la voluntad libre y el autocontrol	67
Capítulo IV La Actitud Intencional, el adaptacionismo y la memética	71
IV.1 Acerca de la intencionalidad	71
IV.2 La Actitud Intencional	75
IV.3 Defensa del adaptacionismo	82
IV.4 La teoría memética	89
Conclusiones	97
Bibliografía	101

## Introducción

Para Daniel Dennett, la mente y sus fenómenos (ego, conciencia, intencionalidad), deben ser estudiados desde la ciencia empírica, por lo que uno de sus reclamos constantes es que no hay una buena razón para creer que la ciencia es incapaz de dar cuenta de tales fenómenos; después de todo, nos dice Dennett, ella ha dado los pasos más importantes respecto de la comprensión del mundo que tenemos los humanos; esto es, ha explicado el sistema solar, la deriva continental, los mecanismos y sistemas corporales de los que no se tenían sino buenas intuiciones e, incluso, le ha descubierto al hombre su origen ancestral y su parentesco con el resto de las criaturas vivientes.<sup>1</sup>

Esto no quiere decir que para Dennett la filosofía sea incapaz de rendir frutos, por el contrario; aunque no tenga todas las respuestas –igual que todo conocimiento humano–, la filosofía puede ser una buena guía acerca de cómo abordar desde la ciencia algunos de los problemas más difíciles con los que se enfrenta.

Así, la concepción que tenemos de la mente es sumamente distinta de la que se tenía incluso a principios del siglo XX, y esto es *thanks very little to progress in philosophy and very much to progress in science*. (Dennett, 2005: 16).

Hemos de seguir, de acuerdo con nuestro autor, el camino de la ciencia para el estudio de la mente por buenas razones, aunque una teoría de este tipo (científica) no ha sido establecida aún. ¿Cómo comenzar, entonces? Dennett asume un papel “modesto” para la filosofía, que se refleja cuando se inspira en la

---

<sup>1</sup> La noción de revolución científica es, aparentemente, apoyada por nuestro autor, aunque hemos de decir que tal idea –que existan revoluciones científicas– ha sido puesta en duda por diversos autores, como Peter Bowler y Steven Shapin, entre otros. Tal discusión queda fuera de los propósitos de este trabajo.

actitud de John Locke en su *Essay Concerning Human Understanding*, de 1690, en la *Epistle To The Reader*:

[I]n an age that produces such masters as the great Huygenius and the incomparable Mr. Newton [...] it is ambition enough to be employed as an under-labourer in clearing the ground a little, and removing some of the rubbish that lies in the way to knowledge; —which certainly had been very much more advanced in the world, if the endeavours of ingenious and industrious men had not been much cumbered with the learned but frivolous use of uncouth, affected, or unintelligible terms, introduced into the sciences, and there made an art of, to that degree that Philosophy, which is nothing but the true knowledge of things, was thought unfit or incapable to be brought into well-bred company and polite conversation.

Para Dennett, en consecuencia, la ciencia que se encargará del estudio de la mente será la ciencia cognitiva, que aun no tiene a su Newton, y el papel de la filosofía ha de ser el mencionado por Locke: desbrozar el camino de aquélla.

Aunque sabemos que difícilmente habrá una postura que cuente “la historia completa” acerca de la mente y del ego ante la especialización en el estudio fisiológico del sistema nervioso y del cerebro, los modelos matemáticos de los procesos de aprendizaje, la lógica de los conceptos mentalistas (creer, saber, desear, etc.) o la historia evolutiva concreta de tales fenómenos, sí habría una postura que puede establecer por dónde empezar, como lo vemos en otro párrafo que refleja muy bien la estrategia del profesor de Tufts:

I do not try [...] to tell the whole story, but to set out the conceptual background against which the whole story must be told, to determine the constraints which any satisfactory theory must evolve. (Dennett, 1991: 156).

Y, para empezar aclarando presupuestos metafísicos, Dennett divide su estudio de la mente en dos partes, en espera de que la ciencia cognitiva avance y –tal vez– halle a su Newton (esta postura realista puede resultar problemática, como veremos). Una parte, pues, consiste en penetrar en la característica de la mente que es más resistente a su inserción en una teoría científica: la conciencia, con

sus cualidades fenoménicas<sup>2</sup>, e introducirla en los procesos físicos y químicos de otro objeto sumamente complejo (aunque no tanto como para ser clasificable de raíz): el cerebro humano.

La otra parte es una teoría del contenido que trata principalmente de cómo hablar de la mente, de cómo podemos clasificar ese fenómeno tan complejo y sus términos relacionados; y este 'cómo hablar de la mente' lo propone Dennett en un intento de legitimar no sólo la psicología popular y los términos mentalistas, sino también al adaptacionismo: la Actitud Intencional.

Esta división en dos partes del estudio de la mente aparece desde *Content and Consciousness* (1969), se desarrolla durante más de tres décadas para producir *Sweet Dreams* (2005) y continúa actualmente.

Para hablar de las teorías del ego y de la conciencia del americano haré lo siguiente: expondré, primero, las bases de su teoría de la conciencia, dejando de lado temas como la heterofenomenología, la discusión de los *qualia* y el funcionalismo, entre otras; esto con el fin de contribuir a notar con mayor claridad las bases de su teoría sobre el yo; al respecto, sostendré que aquello que la define, en último término –y, además de lo mencionado respecto de despejar el camino conceptual para la ciencia–, no es la distinción *hardware-software* (veremos cómo el propio autor insiste en que se trata de una familia de metáforas), sino cierta interpretación del evolucionismo. Allí también encontraremos lo que considero su principal problema: el énfasis que hace en que la continuidad evolutiva entre naturaleza y cultura depende de y se revela con las unidades de transmisión cultural o *memes*.

---

<sup>2</sup> Cualidades subjetivas, similares a los datos de los sentidos, pero con carácter presuntamente universal; por ejemplo, llamamos por convención 'verde' al color verde, pero la *sensación* de verde es privada. Uno de los más famosos ejemplos de la supuesta impenetrabilidad de las cualidades fenoménicas es el ensayo de 1974 de T. Nagel, *What is it like to be a bat?*; Dennett, por su parte, hace un análisis de este concepto en varios de sus libros, haciendo notar que el concepto es vago y conduce al solipsismo (Dennett, 1991, 1996, 2005). No entraré aquí a la discusión acerca de las cualidades fenoménicas o *qualia*.

Esto nos llevará a lo que Dennett llama la Actitud Intencional. Al querer destacar las bases evolucionistas de sus teorías del ego y de la conciencia, tenemos que encarar el problema de la continuidad evolutiva, que va desde los organismos más simples hasta el ego y la conciencia humanos como una rama más del árbol de la evolución. Al respecto, al discutir temas evolucionistas, los argumentos de Dennett se centran en el adaptacionismo, entendido aquí como el tipo de evolucionismo que otorga un peso central, en la configuración de sus rasgos actuales, a la adaptación de las especies al ambiente. El adaptacionismo, por lo tanto, es visto como la fuerza central del mecanismo de la evolución por medio de la selección natural. En efecto, para Dennett, el adaptacionismo que propone se justifica, precisamente, desde las hipótesis de la Actitud Intencional.

A partir de esta tentativa clarificadora, procuraremos analizar lo que Dennett considera como sus aportaciones más importantes a la filosofía, respecto de la mente (Dennett, 2008)<sup>3</sup>, a saber:

- 1.- La idea de que la conciencia no es ningún punto central;
- 2.- La propuesta de que la Actitud Intencional es una poderosa herramienta para predecir y explicar conductas, legitimando lo que hacemos cotidianamente –tratar a otros como agentes–, y guiar las investigaciones empíricas, siempre y cuando se la utilice con cuidado; además, la afirmación de que hay una especie de gradualismo entre los sistemas que se pueden predecir y explicar con la Actitud Intencional.

Al analizar el primer punto, pondremos de relieve su crítica del Teatro Cartesiano, para a nuestra vez poner a prueba y criticar esta tentativa dennettiana. Se hará notar posteriormente que, para hablar de la conciencia y del ego, el norteamericano procede siguiendo tres pasos: el primero es presentar *una familia*

---

<sup>3</sup> Entrevista a Daniel Dennett, recurso electrónico.

*de metáforas* que no deben tomarse en sentido literal, pero que nos ayudarán a pensar en la mente y en sus fenómenos; en segundo lugar, tenemos que las metáforas se apoyan en ejemplos concretos de la ciencia empírica, a veces mostrando los presupuestos metafísicos de ésta y, finalmente, el tercero consiste en el propio paso que muestra el evolucionismo de nuestro autor.

Destacaremos, para finalizar, que las explicaciones de Dennett tienen como punto de apoyo, en última instancia, a la Actitud Intencional, una herramienta ontológicamente poco comprometida con los fenómenos que trata, pero usualmente útil en contextos cotidianos y científicos.

Todo esto nos ayudará, esperamos, a mostrar cómo el diálogo entre ciencia cognitiva, evolucionismo y filosofía de la mente puede ser fructífero para enfatizar algunos aspectos del estudio del ego, como el intento por analizar si éste es un fenómeno susceptible de ser analizado desde una perspectiva diacrónica evolucionista (y que nuestro autor no es el único en señalar, claro) y, a la vez, analizar algunos presupuestos metafísicos que parecen persistir no sólo en el sentido común, sino también en las neurociencias.<sup>4</sup> Creemos que la propuesta de Dennett es fructífera en tal sentido.

---

<sup>4</sup> Intentaremos mostrar esto cuando revisemos los experimentos de Libet, pues creemos que nos pueden ayudar no sólo a mostrar un aspecto del diálogo entre las ciencias cognitivas y la filosofía de la mente que propone Dennett, que consiste en analizar los presupuestos metafísicos de la primera, sino también para señalar que nuestro autor no ataca a “hombres de paja” cuando señala que el ego es una narración y no un punto central.

## Capítulo I

### *La teoría de la conciencia*

Como mencionamos en la introducción, para Dennett la filosofía de la mente es inevitable para la ciencia y viceversa. La estrategia tradicional respecto de esto, nos dice nuestro autor, ha sido tratar de “llenar los huecos” de las preguntas más frecuentes respecto de la conciencia, con los logros científicos en la materia –como se intentará mostrar cuando hablemos del Materialismo Cartesiano– lo cual sería un error.

No se trata, nos dice Dennett, de responder a las preguntas tradicionales desde la ciencia, por el contrario, debemos de poner toda la atención a los datos empíricos y a los modelos de funcionamiento cerebral y, a partir de ahí, poner en duda los supuestos tradicionales de la filosofía, pero también analizar los presupuestos metafísicos envueltos en la ciencia empírica. Y uno de los más importantes presupuestos nos viene del dualismo cartesiano. En la dinámica dennettiana, debemos entonces partir de la perspectiva tradicional acerca de la conciencia: el Teatro Cartesiano, que está lejos de ser una postura olvidada.

### *El Teatro Cartesiano*

De acuerdo con René Descartes, como sabemos, el hombre es una sustancia, una cosa pensante, cuya manifestación se nos da de diversas maneras: en la imaginación, el sentimiento, la voluntad, incluso en el sueño, pero todas remiten a esa sustancia mental. A través del uso de la duda metódica, que parte de poner en suspenso la realidad de cualquier contenido mental, llega a la conclusión de que si se duda, se existe, con lo cual hay por lo menos la certeza de un contenido

mental, aunque este no corresponda –como en el sueño– a lo que acontece en ese momento en la realidad. Nos dice el gran teórico del dualismo:

Pendant que je voulais ainsi penser que tout était faux, il fallait nécessairement que moi, qui le penóais, fusse quelque chose. Et remarquant que cette vérité : *je pense, donc je suis* était si ferme et si assurée que toutes les plus extravagantes suppositions des sceptiques n'étaient pas capables de l'ébranler, je jugeai que je pouvais la recevoir, sans scrupule, pour le premier principe de la philosophie que je cherchais (Discours de la Méthode: IV partie).

Ahora bien, existen, además, la cosas materiales, por lo menos en cuanto las percibimos con claridad y distinción como partes de la sustancia extensa. Para Descartes, pues, el hombre está compuesto de dos sustancias irreducibles entre sí: mente (*res cogitans*) y cuerpo (*res extensa*), las que, sin embargo, se comunican, ya que la primera puede regir hasta cierto punto los destinos de esta parte de la *res extensa* que es el cuerpo humano. Pero el francés, como debemos recordar, tuvo dificultades para explicar sin contradicciones cómo interactúan estas dos sustancias: situó el punto de interacción entre ellas en la glándula pineal del cerebro humano (no de cualquier cerebro), lo que es sin duda una simple invención de implicaciones metafísicas.

En efecto, la debilidad de esta solución al problema de la relación mente-cuerpo será el punto de partida de discusiones posteriores y, en nuestro caso, del análisis de presupuestos metafísicos que persisten en la ciencia cognitiva actual y el intento por explicar a la conciencia y al ego.

El Teatro Cartesiano (TC en adelante) es, de acuerdo con el profesor de Tufts, el escaparate que reúne a todos los elementos percibidos e imaginados en la conciencia. Podemos adelantar que el trabajo que hasta ahora se ha atribuido a este “espacio”, deberá ser distribuido entre distintas partes del cerebro que no ostentan ningún rango especial.

Así pues, esta perspectiva, que ha llegado a ser parte del sentido común, consiste en suponer que hay un lugar central en el cerebro, un proyector de apariencias

donde alguna especie homúnculo o espíritu aprecia, cataloga y valora los contenidos de la experiencia; un punto de vista o proceso independiente de observación dentro del cerebro. Las intuiciones que tenemos, nuestro flujo de conciencia que parece constar de eventos secuenciales, sugieren este lugar central, donde los vehículos de contenidos viajan en una especie de línea causal, y su orden de llegada al TC es el orden en el que se hacen conscientes.

### ***El Modelo del Teatro Cartesiano a prueba***

Como primer punto debemos notar que el TC implica una regresión al infinito: si ese observador interno juzga lo que se le presenta, ¿es que hay dentro del observador otro observador que observa y sopesa lo que el primer observador cree percibir? ¿Y qué hay de ese observador, cómo vigila? ¿O si existen tales “ojos de la mente” que aprecian lo que los ojos de la cabeza le presentan?, ¿es que hay unos ojos de la mente dentro de los ojos de la mente que aprecian las imágenes que les dan los ojos de la cabeza? ¿No es ésta una regresión infinita de imágenes y de espectadores?

El error de esta postura, sostenida por el sentido común, radica –nos dice el americano– en que no se hace ninguna diferencia entre la forma reflexiva de pensamiento y la no reflexiva, y se identifica un estado consciente con el tener meta-creencias que implican el TC o lugar central para el observador<sup>5</sup>. No es que aprehendamos primero la experiencia en el TC para luego construir testimonios o meta-pensamientos o meta-creencias, sino que ese mismo ser capaz de decir lo

---

<sup>5</sup> Es importante tener en mente esta distinción entre la forma reflexiva y la no reflexiva, para apartados posteriores, pues *parece* que en la mayoría de sus escritos el americano la olvida, como cuando habla de un punto crucial respecto de la distinción entre conciencia humana y conciencia como estado de intensa vigilancia o de intencionalidad fuerte. A lo largo de sus publicaciones, Dennett habla simplemente de “conciencia”, entendiendo por tal término un amplio rango de “facultades”, por decirlo de algún modo, que va desde los estados de alerta que se dan en otras especies además de la humana, hasta los metapensamientos, propiamente humanos. Por esta razón, nosotros hablaremos de ‘conciencia humana’, cuando hagamos referencia a la capacidad reflexiva típicamente humana y de ‘conciencia’ cuando hablemos de estados de alerta o de autocontrol.

que se siente ser *x* o estar en la situación *z* es la base de las meta-creencias o meta-pensamientos.

La perspectiva del TC invierte el proceso, pues los pensamientos o creencias, nos dice nuestro autor, no se presentan primero para luego narrarse, sino que se “narran” (se fija el contenido de la percepción) y luego se “presentan” (se articula la expresión):

It is not that first one goes into a higher-order state of self-observation, creating a higher-order thought, so that one can then report the lower order thought by expressing the higher-order thought. It is rather that the second-order state (the better-informed state) comes to be created by the very process of framing the report. We don't first apprehend our experience in the Cartesian Theater and then, on the basis of that acquired knowledge, have the ability to frame reports to express; our being able to say what it is like is the basis for our "higher-order beliefs" (Dennett, 1991: 315).

Esta perspectiva del jefe interno u observador es, sin embargo, una poderosa herramienta para organizar y controlar nuestros estados internos.

Dennett, no obstante, conserva algunos puntos de la imagen tradicional del TC, como la idea de que ser consciente de algo es tener un estado *informado* o la referencia de que existirían estados de primer orden a sí mismo, lo que evitaría caer en la regresión al infinito que implica la creencia en un observador de tales estados.

Entonces, ¿debemos, de acuerdo con el profesor de Tufts, eliminar la idea del TC?, no necesariamente, sólo la creencia en su realidad ontológica: el TC puede ser tomado *metafóricamente*, como un espacio lógico, para expresar “el lugar” donde se encuentran las sensaciones fenoménicas, así como las creencias y los deseos, nos dice nuestro autor. Analizaremos cómo Dennett habla de las creencias y de los deseos (términos mentalistas), cuando lleguemos a lo que el americano llama la Actitud Intencional (*Intentional Stance* [AI, en lo sucesivo]).

Pero podríamos preguntar a Dennett lo siguiente: ¿no es cierto que el dualismo cartesiano o TC es una teoría ignorada, dejada de lado por la ciencia actual? ¿Por qué insistir en analizar un presupuesto que se ha mostrado poco fértil, incluso caduco? ¿Alguien sostiene –o sostendría– en la actualidad, que existe el TC?, o por decirlo de otro modo, se sigue sosteniendo que el ego y la conciencia son origen de, por ejemplo, la voluntad libre?<sup>6</sup> Para esto hablaremos un poco acerca del modo en el que experimentos como el del fenómeno *phi* muestran la creencia en la realidad ontológica del TC y su transformación; es decir, analizaremos el Materialismo Cartesiano.

El Materialismo Cartesiano (MC en adelante) es, de acuerdo con el americano, la tesis que sostiene que hay una línea de meta o punto central en el cerebro donde se hacen conscientes los datos de la percepción que ahí llegan, algo así como la glándula pineal de Descartes, *sin que ello comporte el contacto con la res cogitans*.

El MC es, nos parece, indisoluble del TC, pues el primero es una transformación del segundo, una noción que persiste incluso en las neurociencias, como se verá con los experimentos de Libet. Buscando mayor claridad, me referiré a ambos como “modelo del TC”.

Estamos con Dennett cuando señala que la mayoría de los teóricos de las neurociencias, aunque afirmen que el dualismo cartesiano es falso, no manifiestan que ello implica negar un lugar central en el cerebro donde los contenidos de la percepción se hacen conscientes y que el orden de llegada de la información no es el orden de la experiencia; esto es, no se ha admitido como falso el MC. Se abundará en éste cuando lleguemos a los experimentos de Libet (1985), que no sólo muestran con más claridad los presupuestos del MC (punto central en el cerebro), sino que nos ayudarán a señalar cómo éstos permean las teorías del

---

<sup>6</sup> Abundaremos en este punto a lo largo del trabajo, pero sobretodo cuando hablemos de los experimentos de Libet.

ego y de la conciencia tradicionales, pero también las de nuestro autor: es ahí donde, a nuestro parecer, se encuentra la importancia de la propuesta dennettiana.

Pues bien, un primer acercamiento práctico a las fallas del modelo tradicional de la conciencia de los modelos del TC y el dennettiano, que intenta mostrarse como un principio de economía de mayor poder explicativo, se mostrará con algunos de los siguientes experimentos –bastante simples– y muestran que Dennett no está atacando a un “hombre de paja” al hablar de su –molesta pero inevitable– persistencia, tanto en el sentido común como en la ciencia<sup>7</sup>. El primero de ellos es el *fenómeno phi*.

Este efecto visual consiste, básicamente, en “engañar” al agente acerca de su percepción. Estudiar cómo la percepción nos engaña no es nuevo; recordemos el bastón que al entrar en el agua se “torcía”, como señalaba Descartes, para “enderezarse” apenas se lo sacaba. Pero el efecto phi, tiene relación tanto con el arte como con la psicología. Este efecto se utiliza desde el nacimiento del cine donde una rápida sucesión de imágenes se crea la ilusión de movimiento en el espectador. El ejemplo más sencillo es cuando dos focos separados a no más de 4 grados de ángulo visual se encienden intermitentemente: la ilusión será que un único punto de luz se mueve hacia adelante y hacia atrás o hacia arriba y hacia abajo. Una variante es cuando uno de los focos es rojo y el otro verde: entonces ambos parecerán ser un solo haz de luz que cambia de un color a otro.

Las explicaciones científicas “estándar” reflejan la postura tradicional que, de ser cierta, podría responder a la siguiente pregunta: ¿el engaño o la ilusión del fenómeno phi se crea en la percepción (sentidos) o en la memoria, una vez que la

---

<sup>7</sup> La relación entre sentido común y la ciencia no ha de ser aclarada aquí. Señalemos, sin embargo, que entenderemos por ‘sentido común’ algo así como conocimiento común o actitud ingenua, con lo que se alude al conjunto de opiniones de que dispone la “gente común” para explicarse fenómenos físicos, situaciones de la vida, relaciones interpersonales, etc., sin tener que fundamentarlas en razones científicas (esta es la relación con la psicología popular y las discusiones que le rodean).

información sensorial es procesada por el cerebro? Si el modelo del TC fuera la metáfora correcta con la que nos podemos referir a la conciencia, es decir, si efectivamente hubiera un punto central a donde llegara la información, podríamos decir con facilidad por cuál de las dos alternativas nos inclinaríamos. Pero esto no es así. Para profundizar al respecto, veamos los ejemplos con las revisiones de la memoria, divididas por Dennett en estalinianas y orwellianas.

Las revisiones estalinianas, cuyo nombre surge “en honor” de la dictadura de José Estalín, en la que hechos presentes eran modificados mostrando pruebas alteradas, incluyendo testigos y confesiones falsas. Se cree tener una experiencia de cierto tipo pero en algún punto de su percepción, se falsean las pruebas. Se trataría de una revisión pre-experiencial (*Pre-experiential* [Dennett, 1991: 123]).

Las revisiones orwellianas, por otro lado, son post-experienciales (*post-experiential*), donde al espectador le *parece* tener un recuerdo de sucesos que nunca han ocurrido, como en la novela de George Orwell, *1984*, donde el pasado ha sido modificado debido a intereses políticos. Este tipo de revisión de la memoria deja ver uno de los aspectos fundamentales de nuestras creencias, la distinción entre la apariencia y la realidad. Ese es su principal atractivo: ¿pasó esto realmente o sólo fue una ilusión? ¿Lo que creo tiene fundamentos o es producto de mis fantasías (o el engaño de otros)?

¿Cómo podríamos ver que a escalas de tiempo microscópicas (como las que se dan en los procesos cerebrales)<sup>8</sup> esta distinción no funciona?; acabamos de mencionar el fenómeno *phi* de los colores: la revisión estaliniana diría que los ojos están engañando a la mente (punto central), pues perciben un espectro de color – rojo-pasando-a-verde– que realmente no está “ahí afuera”. Estos experimentos de que los sentidos nos engañan son bien conocidos (como el efecto de refracción de

---

<sup>8</sup> Por ejemplo: saying "one. Mississippi" 1000msec; unmyelinated fiber, fingertip to brain 500msec; a 90 mph fastball travels the 60.6 feet to home plate 458msec; speaking a syllable 200msec; starting and stopping a stopwatch 175msec; a frame of motion picture film 42msec; a frame of television 33msec; fast (myelinated) fiber, fingertip to brain 20msec; the basic cycle time of a neuron 10msec; the basic cycle time of a personal computer 0001msec (Dennett, 1991: 103).

la luz en el vaso con agua y la vara torcida de Descartes). Entonces, la ilusión se da a la par de la percepción.

La revisión orwelliana trataría de dar cuenta del fenómeno más o menos de la siguiente manera: en el momento en que los ojos percibieron ambos focos prendiéndose alternativamente (esto tal vez fue una percepción inconsciente), se activó el ciclo de vías de entrada hasta el punto central del cerebro donde tales vehículos se hacen conscientes, fue entonces que el cerebro “rellenó” la percepción y creó un contenido: el rojo-pasando-a-verde. La idea básica es suponer que el cerebro “pone” algo cuando la percepción real es incompleta: esta es la idea que Dennett llama repleción (*filling in*).

Entonces, en el caso de la revisión orwelliana, primero fue la percepción, luego la ilusión, y habría un “desfase” entre los dos acontecimientos. Se tiene una revisión post-experiencial. *To put it loosely, in these cases is your memory playing tricks on you, or are just your eyes playing tricks on you?* (Dennett, 1991: 123-124).

Si tenemos estos dos tipos de explicaciones, ¿cuál es la que da cuenta de lo que pasa *realmente*? Si el modelo del TC fuera verdadero, sin duda habría una respuesta, pero la evidencia no permite distinguirlo. Ambas revisiones son consistentes con cualquier cosa que el sujeto piense, diga o recuerde, pero el principal error de esta idea es que el cerebro, en vez de rellenar o poner algo, omite o ignora cosas después de “filtrar” los contenidos.

Como ya habíamos mencionado, el modelo del TC es admitido, entre otras cosas, porque sostiene la distinción clave entre apariencia y realidad, pues soy “yo” (el punto central en el cerebro) quien tiene una posición privilegiada para decidir si algo es un engaño o no; además, soy yo quien carga con la responsabilidad moral de mis actos, y no una serie de procesos mecánicos, soy el *agente* dentro del

cuerpo.<sup>9</sup>

La diferencia entre revisiones orwellianas y estalinianas se desvanece, entonces, conforme el tiempo a medir se reduce; se apoya así la hipótesis de que el lugar central o línea de meta de las percepciones o pensamientos (lugar de la conciencia) es una extrapolación geométrica, por lo que según nuestro autor, deberíamos desechar tal hipótesis.

Hay otro fenómeno que muestra lo mismo, se trata del fenómeno del metacontraste:

if a stimulus is flashed briefly on a screen (for, say, 30msec — about as long as a single frame of television) and then immediately followed by a second "masking" stimulus, subjects report seeing only the second stimulus. The first stimulus might be a colored disc and the second stimulus a colored ring that fits closely outside the space where the disc was displayed. People can nevertheless do much better than chance if required to guess whether there were one or two stimuli. (Dennett, 1991: 141).

Las maneras de explicar este fenómeno desde la perspectiva tradicional del materialismo cartesiano son, nuevamente, las revisiones estalinianas u orwellianas: en el primer caso el segundo estímulo impide la experiencia consciente del primer estímulo; aquél le “sale al paso” a éste y termina primero el recorrido hacia el punto central, además de que el primer estímulo se manifiesta de manera inconsciente en las respuestas de los sujetos; en el segundo caso los sujetos sí son conscientes del primer estímulo, pero en la memoria el segundo estímulo reemplaza al primero casi por completo (por eso los sujetos mencionan que hay dos estímulos). Ninguna de las dos posturas puede dar un resultado experimental que incline la balanza hacia un lado u otro (Dennett, 1991: 440).

Estos experimentos muestran un primer acercamiento a la idea principal de los modelos de la conciencia de Dennett: no hay un punto central donde algo se haga

---

<sup>9</sup> No veremos aquí las preocupaciones acerca de la agencia y de la voluntad libre como no sea de paso, cuando veamos los experimentos de Libet; aquéllas son tratadas por el americano en *Elbow Room* (1984) y *Freedom Evolves* (2003).

consciente, sino que regiones del cerebro reconocen sus alertas y se disparan (si pasa  $x$ , entonces haces  $y$ ) sin que el tiempo de los representantes (lo percibido) sea el tiempo de lo representado (lo experimentado de manera consciente): el punto de la medición temporal a escalas microscópicas es uno de los aspectos importantes de los experimentos de Libet, como veremos más adelante, de modo que habría que tenerlo en mente.

Dennett sostiene una cierta teoría de la percepción de la tercera persona –como la ciencia empírica– a lo largo de todo su trabajo y la utiliza para sus modelos de la conciencia, y nos ayudaría a dar cuenta, por ejemplo, del fenómeno phi: la captación de los diferentes elementos en el mundo no es “directa”; es decir, no experimentamos lo que está en el mundo sino hasta después de una serie de procesos de interpretación de las diferentes regiones cerebrales y los órganos de los sentidos (lo que experimentamos no son las ondas electromagnéticas que reciben nuestras retinas, ni lo que recibe la superficie de nuestra piel como respuestas galvánicas, sino un “informe” de tales sensaciones más o menos caótico, como veremos más adelante, cuando hablemos de los modelos de Dennett para la conciencia). Lo central en la percepción es abastecer de elementos relativamente burdos que luego son cotejados, revisados, amplificados o incluso *falseados* por el cerebro, el “cuartel general” que no necesita de otro cuartel general interno (Dennett, 1991: 106).

Ahora bien, ¿cómo explican los modelos dennettianos este fenómeno? Además de lo ya dicho acerca de la teoría de la percepción que el profesor de Tufts sostiene, añadimos que la información sobre el primer estímulo ocupó una posición *funcional* durante un breve espacio de tiempo, aunque no debemos suponer o preguntarnos si fue o no algo que llegó a la conciencia o fue sobrescrito por el segundo estímulo en la conciencia, sino que se integró momentáneamente al flujo de contenidos para luego desaparecer, aunque el por qué de tal desaparición sea un capítulo pendiente hasta la fecha.

El punto es señalar que *parecemos* tener absoluta autoridad sobre los *contenidos* de la mente, pero ¿qué pasa con las ilusiones ópticas o cognitivas?: pareciera que sólo desde el punto de vista de la tercera persona podemos notar estos errores.

¿Podría ser, entonces, que la conciencia en vez de un subsistema (TC) fuera un modo de acción?, es decir, ¿podría ser no un lugar central sino un proceso en el que, cuando la mayor parte de, por ejemplo el córtex, se activara, diera lugar al proceso consciente? No sería el caso, de acuerdo con Dennett, porque sería suponer el modelo del TC de una manera distinta, modelo que equivaldría a decir que la organización temporal de los procesos “en el exterior” dicta los patrones para la secuencia subjetiva o “interior”, a la manera de una pantalla que refleja los eventos del mundo, como el modelo del TC sugiere.

Tal parece que el cerebro no reacciona como un punto focal para producir conciencia, sino que diferentes regiones actúan en diversos momentos en un proceso no paralelo al exterior como lo muestran los sistemas de producción del habla, donde la emergencia de la expresión es precisamente lo que crea o fija el contenido del pensamiento de orden superior expresado y no una región fija, como veremos con el modelo del pandemónium.

Otra de las claves de que el modelo tradicional se mantenga en boga se debe, nos dice el americano, a su comodidad, a que es en gran parte una extrapolación geométrica: suponer que los supuestos tradicionales funcionan para toda situación y escala, o, en este caso, en suponer que los principios que funcionan bien en situaciones normales (macroscópicas) lo seguirían haciendo respecto de la conciencia en la escala microscópica de tiempo de los estados cerebrales.

Por ejemplo, sabemos que la luz de una estrella llega de lugar y tiempo distinto al que percibimos (por la distancia y la gravedad que la atrae en su trayectoria); es decir, el lugar de donde parece venir y el tiempo en que la percibimos depende del

punto de vista del observador. Y esa es una característica cardinal: siempre que hay una mente consciente hay un punto de vista, el punto de vista del observador.

Pretender localizar ese punto de vista dentro del cerebro es un error fundamental, extrapolar este principio aplicable en ejemplos como el de la luz es el supuesto falso: así como encontramos que del punto de vista del observador (cuerpo) dependen ciertas maneras de percibir, hay un punto de vista *hacia adentro*, en la conciencia, al que llegan las percepciones como a una línea de meta.

El norteamericano nos dice que podemos minar la confianza en la perspectiva de la primera persona –donde el modelo del TC encuentra su fundamento– de un modo similar al que Hume socavó el de causalidad, señalando que no hay una relación de necesidad entre lo que consideramos causa y efecto, sino que tal inferencia proviene de la costumbre (Dennett, 1991: 133). Con la introspección ocurre algo similar: queremos decir lo que pasa dentro de nuestros cerebros, lo que hace que surjan hipótesis acerca de lo que sucede, sin que haya una conexión que no sea la de la costumbre. Es decir, nuestras conclusiones no nos vienen del mundo, no las extraemos de él por necesidad ontológica, sino que las creamos.

El modelo del TC implica también algo así como un jefe interno o *Significador Central* (*Central Meaner*), que controla la comunicación con el mundo exterior (Dennett, 1991: 231). Es él quien decide qué palabras han de ser proferidas finalmente –como producto de la idea de que hay un punto central en el cerebro donde las cosas se hacen conscientes– pero contribuye a que surjan problemas: ¿mediante qué proceso el Significador Central elige o articula las palabras o frases que han de ser proferidas? ¿Hay otro significador que así se lo indica? ¿Si allí se encontrara la fuente del significado, en qué idioma se pre-configurarían las palabras antes de salir al exterior? En efecto, si el sentido de las oraciones ya estuviera dado para el Significador Central en ese momento, lo más difícil estaría hecho; pero es difícil que tal sea el caso.

Además, la idea de un Significador Central muestra nuevamente la regresión al infinito de la metáfora de los “ojos de la mente” del modelo del TC ¿hay otro observador dentro del primer observador? El Significador Central, hemos de notar, es algo así como el tradicional yo o ego, fuente de la agencia, la identidad y organizador de las percepciones.<sup>10</sup>

Hemos visto algunos problemas con el supuesto metafísico del Teatro Cartesiano, que parecen implicar una regresión al infinito: ya sea que el punto central en el cerebro se entienda como una especie de homúnculo, o como “ojos de la mente”, así como los problemas del Materialismo Cartesiano al hallarse ante escalas de tiempo microscópicas, posturas que suponen un punto central o de llegada en el que las percepciones se vuelven conscientes. Tanto TC como MC parecen basarse en extrapolaciones geométricas y confianza en la perspectiva de la primera persona, pero, ¿qué propone Dennett para suplantar el modelo del TC? ¿Cuáles son las primeras limitaciones de la propuesta del profesor de Tufts?

### ***El surgimiento de los modelos de la conciencia en Dennett***

Hasta ahora hemos visto algunos supuestos basados en extrapolaciones y la tradición. Pero, ¿en qué consiste la conciencia humana de acuerdo con el profesor de Tufts? Veamos primero el punto de partida de sus modelos acerca de la conciencia.

Para Dennett los humanos somos una especie de máquina biológica, no hay ninguna diferencia fundamental entre una máquina y un cerebro humano: podríamos decir que la diferencia entre complejidad es sólo de grado y no de tipo:

---

<sup>10</sup> Notemos que desde la perspectiva del modelo del TC, perteneciente al sentido común, no hay una distinción clara entre conciencia y ego; parece que ambos surgen en el mismo instante, ninguno es condición de posibilidad del otro: el ego es a la vez autoconciencia y prueba de su existencia, idea perfectamente compatible con las propuestas cartesianas.

¿no sucede acaso lo mismo cuando comparamos a los humanos en general con los animales no humanos?, las diferencias son siempre de grado.

El cerebro humano, el “cuartel general”, tiene como objetivo guiar al cuerpo que está bajo su mando en la vida, en las condiciones cambiantes que se presentan, almacenar información y organizarla para la supervivencia<sup>11</sup> (en primera instancia), “producir futuro” (Dennett, 1991: 144) anticipando situaciones nocivas para el organismo, así como previendo y buscando las que lo beneficien. El cerebro es una especie de máquina de anticipación cuya función central es controlar la conducta en tiempo real.

Sabemos que en el cerebro hay innumerables procesos automáticos, que aunque no eliminan la posibilidad teórica de un espectador central (Descartes admitía que hay “reflejos”, como cuando doblamos las manos al caer) lo hacen prescindible para toda función cerebral. Es un espacio de trabajo global, sus funciones están en todo el córtex y tal vez en otras regiones. Los límites de todas estas especulaciones son, nos dice Dennett, los límites de la ciencia actual: *It is not surprising that we have no good models, yet, of such multiple functionality* (Dennett, 1991: 273). Notemos que el cerebro, considerado así, tiene como analogía a las computadoras: máquinas con una finalidad que muestran que no hace falta una conciencia humana para actuar en el mundo. En el caso de nuestro autor, la analogía cerebro/computadora es simplemente una metáfora que nos ayuda a pensar fenómenos como los de la conciencia y el ego, pero que posiblemente sea abandonada ante los avances de las neurociencias (Dennett, 1991: 455).<sup>12</sup>

---

<sup>11</sup> Un elemento tan o más importante que la supervivencia de los individuos es la reproducción. Pero el análisis de la relación cerebro-sexo nos alejaría del propósito central de este trabajo.

<sup>12</sup> No entraremos aquí en las discusiones acerca del computacionalismo, tal empresa sobrepasa los límites de este trabajo. Nos limitamos a señalar que aquello que define a la postura de Dennett es el evolucionismo.

Cuando, por ejemplo, las circunstancias requieren juicios temporales de cierta finura –por ejemplo en el caso de discriminar un movimiento de izquierda a derecha– Dennett habla de “discriminadores especializados” (Dennett, 1991: 50, 150, 181) que luego envían la información a otra región para que sea utilizada. Pero si existen esos discriminadores especializados, ¿por qué no utilizarlos todo el tiempo?, ¿en qué casos se utilizan los no-especializados?; ¿podemos suponer que existe un discriminador principal para saber cuándo se utiliza uno y cuándo el otro? Dennett nunca explica esto.

La ciencia, entonces, tendrá que dar cuenta de por qué hay mecanismos diferentes para resolver situaciones en las que se involucra el tiempo con diferente rango para el error. Pero como se trata, presumiblemente, de un problema no sólo empírico, sino conceptual, (“el tiempo”, “la conciencia”, etc.) vemos la necesidad de una teoría de la mente interdisciplinar, pero observamos también el surgimiento de uno de los problemas de la postura del autor de *Sweet Dreams* y que S. J. Gould (Gould y Lewontin, 1979) señala: el poco cuidado en las explicaciones de algunos fenómenos que se antojan muy complejos, siempre argumentando que en el futuro se aclarará la situación, acorde con su realismo. Hemos de tener esto en mente cuando hablemos del adaptacionismo de Dennett.

En ciertos puntos, pues, parece que Dennett añada facultades especiales *ad hoc* para explicar procesos según se requiere, y aun cuando Dennett admite al final de *Consciousness Explained* que su teoría de la conciencia dista mucho de ser completa (Dennett, 1991: 455), es posible que los intentos de explicación presentes resultaran más fructíferos si en vez de confiar en explicaciones mono-causales optáramos por una especie de pluralismo, integrando distintos campos del conocimiento en los esfuerzos por explicar algunos fenómenos; como señala Gould (Gould & Lewontin, 1979).

Volviendo a Dennett: si dejamos de lado el cartesianismo, donde la conciencia y el ego humanos –la sustancia pensante– son inmateriales, y nos olvidamos también

del Materialismo Cartesiano –la idea del punto central del cerebro–, los seres humanos pueden ser vistos desde el ángulo de las teorías funcionalistas, según las cuales se trata de una gran organización de departamentos intercomunicados, regiones corporales realizando determinada(s) función(es), fijándose y/o simplemente logrando objetivos, etc.

Si podía creerse que el cuerpo humano es como una especie de máquina, pero que su conciencia es algo diferente, con las computadoras se ha mostrado –según nuestro autor– que una relación funcional de piezas puede controlar procesos para discriminar, inferir, calcular, anticipar, hallar relaciones complejas entre elementos, y producir conocimiento nuevo de manera muy semejante a como lo hace una mente humana, incluso de modo más eficiente y de más alcance.

En esta línea de pensamiento hardware/software podemos ir “construyendo” al ser humano, ese robot hecho de micro-robots, poco a poco, pieza por pieza, y el primer paso para esto es voltear a ver a las máquinas para *imaginar* cómo es posible el surgimiento de la mente sin necesidad de homúnculos en un teatro, máquinas cuyas partes no necesitan ser conscientes para realizar sus funciones. Ésta es a final de cuentas la tarea que Dennett, entre otros, se impone respecto de la conciencia: imaginar cómo es posible una teoría científica de ella.

Mencionaremos este proceso de manera muy breve, pues lo que queremos destacar es la estrategia de Dennett. Retomemos el ejemplo tan utilizado desde hace decenios de la llamada máquina de Turing. Esta máquina –más bien un modelo abstracto– está inspirada, de cuerdo con nuestro autor<sup>13</sup>, en el proceso del pensamiento humano. El punto básico es descomponer el pensamiento humano en sus elementos básicos, mecánicos, cuando trata de resolver problemas matemáticos y de crear una máquina que pueda hacer la misma cosa. Este es el

---

<sup>13</sup> Este punto puede someterse a escrutinio, debido a que no es claro que Turing haya tenido al proceso de pensamiento humano como “inspiración” para el modelo que lleva su nombre, pero no es crucial para nuestros fines, que se limitan a mostrar los puntos básicos que llevan a Dennett a proponer sus metáforas acerca de la mente.

concepto abstracto que abrió paso a las computadoras, partiendo de la pregunta que se expresa más o menos así: ¿cómo es que un humano realiza en su cerebro operaciones matemáticas? (Dennett, 1991: 212).

Derivada de esta idea de la máquina de Turing, tenemos la máquina de von Neumann, producto de un proceso serial de cálculo deliberado. Ambas máquinas (pues la de von Neumann es una especie de máquina de Turing más compleja) aíslan un tipo de flujo de información mecanizándola (inspiradas en el flujo de la conciencia humana) que, sin embargo, es relativamente plástico debido a la memoria, pues se almacenan distintos programas o datos. Son listas de instrucciones que responden mecánicamente a los patrones preconfigurados.

El siguiente paso son los sistemas de producciones, descendientes de la arquitectura de von Neumann, donde las producciones, que son mecanismos de reconocimiento de patrones o una especie de órganos que funcionan condicionalmente bajo la forma “si, entonces” y que se disparan, se almacenan en la memoria de trabajo. Todas las computadoras actuales son una especie de sistema de producciones con miles de estos sistemas u “órganos” detectores de patrones en espera de percibir a su patrón para actuar.

Estos sistemas resuelven los posibles conflictos no mediante un jefe interno, sino de manera no automática o preconfigurada siguiendo principios básicos debidos al mismo flujo de información, creando una especie de “callejón sin salida” que crea un espacio de problemas que los detectores de patrones van resolviendo por eliminación de acuerdo a su configuración.

De acuerdo con Dennett la conciencia humana es *muy parecida* a una máquina de von Neumann, un procesador serial con contenidos que se suceden y que pasan a través de un cuello de botella en un acumulador (Dennett, 1991: 210). Pero, ¿cuál sería el “cuello de botella del acumulador” del cerebro? Nuevamente nos encontramos con los límites de la explicación dennettiana al respecto, pues su

respuesta es que no hay uno fijo. Si la conciencia de algo requiere de la intencionalidad fuerte (concentrarse en algo, como veremos) pueden ser distintos sistemas o subsistemas cerebrales los que se centren en ese momento en el objetivo. Un esbozo de la interacción entre “circuitos” y cómo se define cuál coalición de ellos está al mando, es un capítulo pendiente para Dennett hasta la fecha.

Pero con las limitaciones de la metáfora *hardware/software*, hacemos notar una de las partes más importantes de la estrategia de Dennett frente al problema de la conciencia humana: ver cómo a partir de modelos más simples se puede crear algo tan complejo como aquella. Es una especie de evolucionismo hacia atrás o “*reverse engineering*” (Dennett, 1995: 212), pues se trata de un intento por develar la arquitectura de la mente humana mediante un producto posterior, entendiendo el asunto solamente como metáfora, como herramienta de pensamiento, una especie de truco utilizado por el americano para comenzar a eliminar la idea del TC; el peligro, habrá que recordarlo, se plantea cuando la metáfora se toma en sentido literal (Dennett, 1991: 455).

### ***Los modelos y las metáforas de la conciencia***

Hemos visto de manera general las consideraciones del americano respecto de la conciencia humana, pero, ¿cuáles son los modelos concretos que propone? Pues bien, para comenzar con esto, haremos notar cómo los modelos del profesor de Tufts van modificándose con el tiempo, revelando lagunas en cuanto a su acercamiento al problema de la explicación de la conciencia. Si bien Dennett no renunciará del todo a la metáfora *hardware/software* que se muestra principalmente en *Consciousness Explained*, las modificaciones sugieren que podríamos renunciar a la familia de metáforas inicial, como intentaremos mostrar, para quedarnos con el intento de naturalización del ego y la conciencia.

Si logramos nuestro objetivo, la situación será tal que mostraremos que el eje del pensamiento de Dennett reside no en las metáforas, sino en el evolucionismo, donde hallaremos su aspecto más interesante respecto del ego y de la conciencia, pero también de lo que podrían ser sus mayores limitaciones en relación a lo que él, considera constituir la continuidad entre evolución biológica y cultural, que depende de las unidades de transmisión cultural o memes, como veremos en el capítulo final de este trabajo.

Comenzaremos con el modelo del pandemónium que, aunque no es un modelo de Dennett, el americano se basa en él para gran parte de lo que propone acerca de la conciencia humana. Este modelo es originalmente de la inteligencia artificial y designa a un sistema organizado donde varias coaliciones de sub-sistemas o partes especializadas compiten por el mando de la situación.

Este método funciona para eliminar “una inteligencia demasiado grande” (Dennett, 1991: 251) como para ser aclarada y las acciones se explican a partir de procesos mecánicos más pequeños y relativamente estúpidos que actúan en conjunto.

Los sistemas en pandemónium implican esfuerzos combinados, algo de caos, “demonios” especialistas y no especialistas, multifuncionales; no hay jerarquías ni modelos prediseñados inamovibles, no se trata de un sistema rígido como en el caso de las teorías homunculares donde hay una organización prácticamente inquebrantable respecto de los medios y fines. El control es usurpado, es oportunista y algo central en él es que carece de diseño (como si dependiera de un producto evolutivo reciente).

Esta idea se refleja muy bien en los actos de habla, por ejemplo con los *lapses* del lenguaje (e incluso con los *lapses* freudianos en general), donde se dice algo que tiene sentido pero que la “mente consciente” no quería decir; de la misma manera otros casos respecto del habla (su producción) sugieren a Dennett que éste es el modelo adecuado para la conciencia humana.

Por otro lado, este modelo es opuesto, claro, al que supone un Significador Central o jefe que ostenta todo el poder. Respecto de los actos del habla, el modelo del Significador Central muestra sus problemas cuando se trata de saber en qué momento se configura un mensaje que será proferido, esto es, ¿en qué lenguaje estaría ese mensaje pre-verbal? Si el mensaje ya está configurado al momento de llegar al Significador, el trabajo difícil ya está hecho y, de ser así, ¿hay un significador dentro del Significador? Vemos nuevamente los problemas de regresión al infinito que implica la hipótesis de suponer que existe algo como el TC.

Basado en un contexto de pandemónium, tenemos el Modelo de las Versiones Múltiples, que ya es propiamente un modelo de Dennett, donde el funcionamiento del cerebro según este tipo metafórico es como sigue –recordemos la teoría de la percepción que Dennett sostiene–: los estímulos sensoriales (visuales, olfativos, etc.) dan lugar a una cadena de procesos de discriminación realizados en distintas regiones del cerebro, que gradualmente aumentan en precisión; también en diferentes regiones cerebrales se registran los datos, se emiten juicios y, en ese momento, los discriminadores –recordemos que los argumentos de Dennett son un tanto vagos en esto, y parece admitir discriminadores según se requiera– se ocupan de localizarlo en el tiempo (inicio y término), “registrando” su forma, color, movimiento, etc. y finalmente reconociendo el objeto.

Parece que en todos los casos el cerebro sigue el mismo proceso para identificar, reconocer y manipular objetos en el mundo; se puede apreciar mejor esta serie de procesos realizados en múltiples regiones cerebrales con los experimentos de Gazzaniga (2008) a pacientes con el cerebro dividido que veremos en el capítulo siguiente. Sin embargo,

The natural but naive question to ask is: Where does it all come together? The answer is: Nowhere. Some of these distributed contentful states soon die out, leaving no further traces. Others do leave traces, on subsequent verbal reports of experience and memory, on "semantic readiness" and other varieties of perceptual set, on emotional

state, behavioral proclivities, and so forth. Some of these effects — for instance, influences on subsequent verbal reports—are at least symptomatic of consciousness. (Dennett, 1991: 134-135).

La consecuencia fundamental de este modelo es que *no* hay un momento preciso en el que algo se haga consciente, no hay un punto en el procesamiento de datos en el cerebro que pueda calificarse de instante de la conciencia; no hay línea divisoria o línea de meta, si se fija una será de manera arbitraria y aquélla se desvanece cuando se consideran procesos microscópicos en el cerebro, como se pudo notar con las revisiones de la memoria (orwellianas y estalinianas) ya discutidas más arriba (Cf. p. xii). El suponer la línea divisoria consciente/inconsciente es suponer una *línea* o *cadena* causal donde fluyen los vehículos (del ojo a la conducta, por ejemplo), cuando se trata en realidad de una *red* causal, con múltiples vías, procesos paralelos continuos, integración y deterioro de contenidos, etc.

No hay un momento preciso en que un determinado evento se vuelve consciente, si está en la memoria, está en la conciencia (Dennett, 1991: 132): con el Modelo de las Versiones Múltiples el acto de “escribir en la memoria” se convierte en el criterio para la conciencia. Somos conscientes de aquellos contenidos que se alojan en nuestra memoria, por breve que sea el lapso en el que lo hacen. Debería sernos suficiente, dice Dennett, con saber que las regiones del cerebro insertan contenidos en el flujo que conduce a la modulación de la conducta, sin postular ninguna propiedad adicional que sea metafísicamente dudosa, como el modelo del TC (Dennett, 1991: 132).

El problema que aquí surge es que el americano jamás desarrolla una teoría de la memoria, apenas asoma una mejoría al respecto cuando habla de su teoría del eco presentada en *Sweet Dreams* (Dennett, 2005), llamado modelo de popularidad o fama cerebral (*Cerebral Celebrity*, Dennett, 2005: 136), que sostiene que un contenido crea la *ilusión* de un mando durante un tiempo al mantenerse en la memoria, punto clave para la teoría funcionalista del americano, pero si ella es el criterio para la conciencia humana, ¿qué estamos entendiendo por ‘memoria’?;

en *Brainstorms* (1979) la memoria es simplemente “M”<sup>14</sup>, y cuando llegamos al *Sweet Dreams* (2005), es la memoria episódica la que juega el papel clave en el modelo de la popularidad o fama cerebral. Vemos, pues, que la metáfora inicial se modifica.

M, según nuestro autor, en *Brainstorms*, es el almacén de las percepciones –que se dan a muchos niveles debido a la complejidad (formas, colores, localización espacial y temporal, etc.)– y los pensamientos, *no es ninguna noción de la psicología* (Dennett, 1981: 158). ¿De qué memoria *podríamos* estar hablando, o con cuál podríamos relacionarla entonces (episódica, procesal, semántica, icónica, etc.)?

Lo que podemos decir es que a Dennett le hace falta una teoría de la memoria aunque, finalmente, como mencionamos, en *Sweet Dreams* (2005) nos habla directamente –sin profundizar demasiado– de ese difícil tema: la memoria episódica es catalogada por Dennett en aquel libro como *uno de nuestros más grandes talentos* (Dennett, 2005: 169) y, nuevamente, el criterio para la conciencia humana; Dennett recurre al mismo cuento (Just So Story) que utiliza en el caso de la evolución de la conciencia –que veremos un poco más adelante– para hablar del surgimiento de tal talento, diciéndonos que es un hábito surgido de la autoestimulación (en todo caso, vale la pena notar hasta este momento que la metáfora del eco en la memoria episódica es sostenida por la teoría evolucionista de corte adaptacionista que el americano defiende).

Esto es, antes de la memoria episódica, talento que permite al cerebro humano ser invadido por el lenguaje, sólo había conciencia y ego de animales no-humanos, después de ella, hubo humanos como complejo cultural.

---

<sup>14</sup> Aunque en *Brainstorms* (1979) habla de memoria de corto plazo y memoria ecoica, el tipo de memoria base de su modelo es M, que no pertenece a las nociones de psicología conocidas hasta ese momento. En todo caso, estaríamos viendo nuevamente la confianza en el futuro de la ciencia que le da su realismo.

Un punto importante a señalar aquí, y que vale la pena recordar por su estrecha relación no sólo con el ego, sino con la postura evolucionista de Dennett y su creencia en la continuidad biológica-cultural, es que la conciencia no escinde el mundo en dos: hay tipos de ella que dependen del curso evolutivo de una especie, pero ¿es la memoria entonces la que escinde el mundo en dos? ¿Es la auto-estimulación que la produjo?; podemos decir que no, pues la conciencia, como el ego, se da en grados<sup>15</sup>, como nos esforzaremos por mostrar más adelante.

Por ejemplo, volviendo a la cultura como punto de inflexión del ego y la conciencia humanos, cuando no *somos* conscientes de algo por falta de entrenamiento, como en el caso de un niño que no alcanza a notar algunos de los gestos faciales de los adultos (en el caso de que no sean claros y dirigidos a él) hasta que está entrenado en ello, como pueden ser la benevolencia, la simpatía, la compasión, la malicia, etc., no somos conscientes de algo porque no hemos sido capaces de entrenar a los mecanismos cerebrales para reconocer tales aspectos.

Otro ejemplo es cuando buscamos algo (unas llaves, un reloj, etc.), y aun cuando el objeto buscado está dentro de nuestro rango de visión, no lo encontramos, no somos conscientes de él, sino hasta que fijamos la atención. Puede notarse que en algunos casos ese “no ser consciente de”, depende del aprendizaje social.

Podría decirse que hace falta *narrabilidad* para que seamos conscientes de algo, es decir, tiene que haber la capacidad de identificar y re-identificar el objeto, examinarlo, asirlo, que sea fruto de muchos movimientos sacádicos. Aquí está nuevamente la relación con la memoria. La conciencia humana desde el esbozo que Dennett propone –y que no es empíricamente exacto como ya hemos mencionado–, surge así:

---

<sup>15</sup> A lo largo de este trabajo hablamos del gradualismo de Dennett sin abordar el problema de las mutaciones o saltos. En todo caso, parece que el americano no tendría problema en admitir saltos evolutivos en, por ejemplo, la aparición del lenguaje en la *Just So Story*, o “cuento” evolucionista que veremos.

Hay muchas redes cerebrales modulares<sup>16</sup> (1) activas en paralelo que procesan información de manera inconsciente [...] una información (2) se vuelve consciente si la población neuronal que la presenta se moviliza por amplificación de atención de arriba abajo (3) en un estado de actividad coherente a nivel cerebral que afecte muchas neuronas distribuidas por distintas áreas del cerebro. La conectividad de larga distancia de esas “neuronas del espacio de trabajo” puede, cuando éstas están activas con una duración mínima, (4) hacer que la información esté disponible para varios procesos, incluidas la evaluación y la acción intencional. Esa disponibilidad general de la información en todo el espacio de trabajo es (5) lo que experimentamos en forma subjetiva como un estado consciente. (Dennett, 2005: 132).

Sin embargo, esto sí deja algo sin explicar: la actividad de cierta estructura neuronal es condición suficiente y necesaria para que haya conciencia humana, ¿y luego qué ocurre?, es decir, no sabemos por qué esa actividad asegura el que dicha estructura permanezca “en el poder” para crear la ilusión de que hay alguien al mando, hay que analizar los poderes funcionales de la captación de la atención. Ésa es la “pregunta difícil” para el profesor de Tufts (Dennett, 2005: 127), sobre la que hay que trabajar de manera interdisciplinaria y, como ya mencionamos al principio, tiene con sus lectores un capítulo pendiente hasta la fecha.

Es importante, sin embargo, señalar que este tipo de reflexividad *no* se refiere a la capacidad típicamente humana de pensar acerca de nuestros pensamientos, sino “simplemente” a la habilidad de *recordar*. Este tipo de reflexividad la tienen incluso los animales no humanos –como los perros– para reconocer a sus dueños aun después de mucho tiempo, o para recordar a alguien que los ha maltratado, recordar olores, etc. Esto implica que, igual que la memoria, la conciencia y el ego, la reflexividad se da en grados. Es decir, si la memoria es el criterio para la

---

<sup>16</sup> Dennett no aclara qué entiende por “redes modulares”. Dada su simpatía por los modelos en pandemonium, podemos suponer que entiende algo como “regiones cerebrales especializadas que cumplen funciones específicas y sólo esas funciones específicas”: *what is being stressed here is only that these [modules] are specialist networks with limited powers of information processing* (Dennett, 2005: 132).

conciencia, entonces el desarrollo de cierto tipo de memoria se convierte en la base del proceso de la reflexividad humana (pensar acerca de nuestros propios pensamientos). El problema es, reiteramos, que Dennett nunca desarrolla una teoría de la memoria. Se limita a señalar que la memoria ecoica es la clave para la conciencia humana. En los humanos esta capacidad es mucho más desarrollada que en el resto de los animales debido, además, a la cultura:

Now perhaps dogs have similar reflective episodes in their inner lives; if they do, then surely they are just as conscious as we are, in every sense. But I hypothesize—this is the empirical going-out-on-a-limb part of my view—that they do not. Events in their echo chambers damp down to nothing after a couple of reverberations, I suspect. Why? Because they do not need such an echo chamber for anything, and it is not a by-product of anything else they need, and it would be expensive (Dennett, 2005: 169).

La conciencia tiene tal vez sus orígenes evolutivos en los estados de intensa vigilancia que vemos en algunos animales no-humanos, como los animales de presa (Dennett, 1991: 180).

El asunto de la memoria es, como dijimos, una laguna clave: Dennett nunca nos habla de la memoria a profundidad y, al ser algo tan importante en su teoría, fundamento de todos los grados de conciencia —desde un ave de presa cazando hasta la cultura humana— uno supondría que habría al menos un bosquejo de teoría, pues si ha de haber una ciencia de la conciencia, deberá entonces existir una ciencia de la memoria (algo que ya se hace, pero que Dennett nunca aborda). El único esbozo que tenemos es lo ya mencionado en relación a los modelos dennettianos, lo que parece insuficiente.

Señalamos, además, que la conciencia es un estado informado, dependiente de lo que Dennett llama relación fuerte de intencionalidad (Dennett, 1991: 346), es decir, que haya concentración suficiente para que un vehículo se transforme en contenido por las coaliciones de regiones cerebrales al mando, por eso mismo el paso de lo consciente a lo inconsciente es maleable, no hay un límite rígido.

A continuación, al mencionar la evolución de la conciencia, hablaremos de cómo esta postura de Dennett da lugar a confusiones acerca de un posible dualismo, para lo que habremos de distinguir entre conciencia refleja e irrefleja. Hablaremos de los memes, unidades de transmisión cultural análogas a los genes de acuerdo con nuestro autor, pero dejaremos su crítica para un apartado posterior. Nos conformaremos, por el momento, con hacer referencia a cómo tales unidades de transmisión cultural han influido según Dennett, en la configuración de la conciencia humana.

### ***La evolución de la conciencia***

No siempre ha habido conciencia, por lo mismo, es plausible pensar que ésta es producto de la evolución biológica, pero el rasgo determinante de la conciencia humana se debe a la evolución cultural: el surgimiento de los límites biológicos dio paso al surgimiento de las razones, al surgimiento de la producción de futuro, a la comunicación, a la auto-estimulación, a la conciencia humana, o “hablar con uno mismo”; necesitamos, para esto, “sólo un cuento” (Just So Story), que refleja, además del optimismo dennettiano respecto del futuro de la ciencia –pues nuevamente nos dice que tendremos que esperar a que la ciencia descubra los detalles (Dennett, 1984: 38)–, su estrategia adaptacionista:

Once upon a time there were creatures who had a full complement of working sense organs “informing” them of conditions in the world, [...] they had coordination problems with other members of their species, and the apparently optimal solutions to these problems required rudimentary forms of “communication” [...] it sometimes happened that when one of these creatures was stymied on a project, it would “ask for help”, and in particular, it would “ask for information.” Sometimes the audience present would respond by “communicating” something that had just the right effects on the inquiring creature [...] causing it “to see” a solution [...] then one fine day an “unintended” short-circuit effect of this new social institution was “noticed” by a creature. It “asked” for help in an inappropriate circumstance, where there was no helpful audience to hear the request and respond. Except itself! When the creature heard its own request, the stimulation provoked just the sort of other-helping utterance production that the

request from another would have caused. And to the creature's delight it found that it had just provoked itself into answering his own question! (Dennett, 1984: 38-40).<sup>17</sup>

La anterior no es una hipótesis descabellada o gratuita, pues el hecho de hacerse preguntas a uno mismo no es una práctica tan inútil como lo sería dejarse propina a uno mismo cada vez que se sirve de comer:

There is considerable evidence, drawn from experiments and from studies of aphasias and other disorders, showing that the processes of speech production and speech comprehension are not mirror-images of one another; in hearing and understanding a sentence, one does not just stuck it in trough the same brain machinery one otherwise uses for formulating and uttering a sentence, only running in reverse. So there is no reason to suppose that the process of formulating, uttering, hearing and comprehending a sentence would simply leave one back where one started cognitively (Dennett, 1984: 41).

Para Dennett la conciencia humana es producto de tres procesos evolutivos superpuestos, que aceleran su constante re-diseño: la evolución genética, la plasticidad fenotípica y la evolución cultural o, dicho de otro modo, en primer término está la selección genética<sup>18</sup>, luego el efecto Baldwin (que veremos enseguida) y al final los memes: efectos secundarios que han dado lugar a la conciencia humana; mediante los memes, nos dice nuestro autor, hemos aprendido a cómo aprender, para después aprender a cómo aprender a aprender, etc.; es decir, al ser portadores de pensamientos y meta-pensamientos el hombre no tiene que comenzar desde el principio, ni por sí solo, a desarrollar buenos trucos.

Aunque sabemos que la naturaleza no es un agente y no puede tener objetivos, con el efecto Baldwin, los más adaptables al medio sobreviven y su plasticidad – en esto consiste el benéfico efecto– lleva a los de sus características a evolucionar en la dirección señalada y así se modifica lentamente el genotipo.

---

<sup>17</sup> Para Dennett, esta hipersimplificación del origen evolutivo de la conciencia humana sería apoyada, refutada o en todo caso puesta a prueba con ayuda de la Actitud Intencional como herramienta de la investigación empírica, como veremos al final. Vale la pena, además, recordar la anotación ya hecha acerca del uso que nuestro autor hace del término 'conciencia' (*Cf. p. x*, nota al pie número 5).

<sup>18</sup> Para esto nuestro autor sigue a Richard Dawkins, quien afirma que los genes, y no los organismos, son objeto de la selección natural.

Este efecto se parece a la idea de que las características adquiridas se heredan (relacionada con Lamarck), pero no es así. La plasticidad fenotípica favorece a los que desarrollen aprendizaje para su supervivencia y, con las condiciones necesarias y regulares a lo largo del tiempo, los que se adapten mejor harán que su genotipo se mueva en la dirección del “buen truco” (Dennett, 1995: 77),

Pero si el efecto Baldwin acelera la adaptación de los organismos mediante adquisición y permanencia del aprendizaje, aparece como una suave presión ante la fuerza de los memes, pues mediante ellos se pueden aprender buenos trucos de humanos que ni siquiera son antepasados genéticos directos (Dennett, 1995: 80). Los memes, por su parte, dependen básicamente del desarrollo del lenguaje (el lenguaje es el meme más importante), la lengua materna es la fuente más poderosa de disciplina para las mentes humanas.

Hacemos notar que para el americano el surgimiento y evolución de los memes obedece *a todas las leyes de la evolución* (aunque a mucha más velocidad), no son metáforas de la transmisión cultural respecto de la biológica (genes), sino su continuación. Por esto, cualquier cosa que se aplique a la primera se aplica a la segunda, pero no al revés.

Así como los genes son transportados y se replican en vehículos para genes (los organismos), los memes son transportados por libros, imágenes, dichos, herramientas, edificios: *a wagon with spoked wheels carries not only grain or freight from place to place; it carries the brilliant idea of a wagon with spoked wheels from mind to mind* (Dennett, 1991: 204).

La conciencia humana –invadida por memes– es una especie de software implementado por la evolución sobre el hardware-cerebro, otra metáfora arriesgada, pero notemos también lo que señala al final de *Consciousness*

*Explained* acerca del uso de metáforas, teniendo en mente el papel para la filosofía que propone cuando habla de Locke:

it is a beginning, because it breaks the spell of the enchanted circle of ideas that made explaining consciousness seem impossible [...] All I have done, really, is to replace one family of metaphors and images with another [...] It's just a war of metaphors, you say — but metaphors are not "just" metaphors; metaphors are the tools of thought. No one can think about consciousness without them, so it is important to equip yourself with the best set of tools available. Look what we have built with our tools. Could you have imagined it without them? (Dennett, 1991: 455).

Al uso de esta familia de metáforas le subyace la idea de evolución de la conciencia, pues como vimos, los modelos del profesor de Tufts van modificándose con el tiempo, quedando lagunas importantes como en el caso de la memoria, sumando y restando detalles con el tiempo, pero la afirmación que no se modifica es la de que la conciencia es un producto de la evolución y que, por lo tanto, no es origen, sino producto.

Recapitulando brevemente: la conciencia humana *puede* ser pensada como pandemónium, como modelo de múltiples versiones, como fama cerebral, como máquina de von Neumann; una visión así es apoyada por pruebas empíricas, pero lo que le es particular, dejando de lado las metáforas, es, como insistimos, que no escinde al mundo *en dos*<sup>19</sup>. ¿Qué significa esto?: la respuesta es que podemos encontrar conciencia en animales no-humanos, no se “enciende ninguna luz” con la conciencia, que se da en la naturaleza en grados.

Pero, ¿no es cierto que para nuestro autor lo *particular* de la conciencia *humana* es que se trata de un software compuesto por memes, lo cual implica nuevamente la metáfora inicial? ¿No parece una especie de dualismo afirmar que hay conciencia humana memética y conciencia animal como estados de alerta? Para contestar esto necesitamos revisar la teoría del ego del americano, esforzándonos por mostrar cómo en ella encontramos más pistas para pensar que las metáforas

---

<sup>19</sup> Con esto no queremos decir que no haya divisiones entre grados, aunque tal vez éstas no sean tan claras como en el dualismo.

pueden ser dejadas de lado, pues el ego es visto como autocontrol, introduciendo nuevamente la idea de continuidad biológica.

Pero es posible que la postura que podría ver en Dennett una especie de dualismo surja porque nuestro autor parece haber olvidado la crítica que hace al sentido común al no diferenciar entre pensamientos y meta-pensamientos, que ya mencionamos al hablar de uno de los errores del modelo del TC (Dennett, 1991: 306): en los escritos del profesor de Tufts jamás se hace explícita la distinción entre conciencia refleja e irrefleja, humana o de otra especie; tal distinción se da por aclarada (es difícil creer que Dennett haga la distinción explícita para luego olvidarla en el momento importante, al hablar de la conciencia humana).

Recordemos lo señalado respecto de la crítica de Dennett al modelo del TC (Cf. p. ix), que no distingue entre conciencia refleja e irrefleja, así como lo dicho en ese mismo apartado, los que nos hacen pensar que la distinción es clara para Dennett; ésta, sin embargo, –y para no confiar en que se trate de un descuido de nuestro autor– es clara cuando habla del ego humano y los distintos grados en que se presenta en la naturaleza<sup>20</sup>, pero más aun, cuando abordemos lo que llama la Actitud Intencional (*Intentional Stance*), y de cómo ésta puede ayudar en la investigación empírica, pues veremos el ejemplo de los monos verdes (vervets) y de los zopilotes aura, y revisaremos los distintos ordenes de pensamiento (recordemos también que Dennett nos habla de la memoria episódica como un proceso de reflexión que no es la capacidad típicamente humana de pensar nuestros propios pensamientos).

Además, y por último, tenemos el párrafo siguiente, que puede guiarnos por el camino que he señalado:

---

<sup>20</sup> Algo similar ocurre cuando nuestro autor habla de la evolución, pues desde los primeros escritos utiliza la palabra ‘*design*’ –diseño–, incluso en *Darwin’s Dangerous Idea*, de 1995, podemos hallar el término en incontables ocasiones, cosa un poco arriesgada, sabiendo que ‘*design*’ tiene una fuerte carga creacionista y normalmente se contrapone a ‘evolución’. Lo mismo pasa con el concepto de ego: sabemos que la distinción está allí sólo cuando leemos ‘ego biológico’.

I have ventured (1) the empirical hypothesis that our capacity to relive or rekindle contentful events is the most important feature of consciousness –indeed, as close to a defining feature of consciousness as we will ever find; and (2) [...] that the Joycean machine in our brains is a virtual machine made of memes. *These are independent claims*. If the meme-hypothesis were roundly defeated by the discovery –the confirmation– of just such echoic systems at play in the brains of nonhuman animals, I would then agree, for that very reason, that the species having those echo-chambers were conscious in just about the way we are –because that’s what I say consciousness is. The price I’d pay for that verdict is the defeat of my bold claim about software and *virtual* machines, but I’d still be getting a bargain, since the other side would be relying on the same theory of consciousness *as a theory of consciousness* in order to establish the relevance –to riddles about consciousness– of their discoveries. (Dennett, 2005: 171-172).

Cuando hablemos del ego, además, trataremos de ver cómo esto nos pudiera ayudar a desligarnos de una teoría que parece dudosa (la teoría de los memes, de la cual hablaremos en el último capítulo, por ahora basta con señalar nuestro propósito), sin sacrificar lo que considero el punto fuerte de la teoría del ego y de la conciencia en Dennett: el esfuerzo por naturalizar al ego buscando analizar y desechar los presupuestos metafísicos que se le relacionan normalmente, desde el sentido común. Que el ego humano se nutra y configure fundamentalmente con y desde la cultura, parece una hipótesis bastante plausible, que la cultura sea los memes, como Dennett cree, es lo que parece difícil.

## Capítulo II

### *La teoría del ego*

Tradicionalmente, se ha entendido al ego humano como el elemento constituyente del individuo consciente, punto focal de los fenómenos físicos y de la identidad personal. Tal vez el ego humano en su forma más conocida es planteado por Descartes, quien lo constituye en punto de partida de su sistema: *pienso luego existo*. El francés lo concibe como sustancia pensante, como entidad permanente y autónoma:

Je suppose donc que toutes les choses que je vois sont fausses; je me persuade que rien n'a jamais été de tout ce que ma mémoire remplie de mensonges me représente; je pense n'avoir aucun sens; je crois que le corps, la figure, l'étendue, le mouvement et le lieu ne sont que des fictions de mon esprit. Qu'est-ce donc qui pourra être estimé véritable? Peut-être rien autre chose, sinon qu'il n'y a rien au monde de certain.

Mais il y a un je ne sais quel trompeur très puissant et très rusé, qui emploie toute son industrie à me tromper toujours. Il n'y a donc point de doute que je suis, s'il me trompe; et qu'il me trompe tant qu'il voudra il ne saurait jamais faire que je ne sois rien, tant que je penserai être quelque chose. De sorte qu'après y avoir bien pensé, et avoir soigneusement examiné toutes choses, enfin il faut conclure, et tenir pour constant que cette proposition: *Je suis, j'existe*, est nécessairement vraie, toutes les fois que je la prononce, ou que je la conçois en mon esprit. (Les Méditations Métaphysiques: Méditation seconde)

Pero con Dennett, el ego humano es expulsado de la conciencia, porque el núcleo *res cogitans* es eliminado: no hay interioridad en el sentido de que sea una especie de recipiente o de teatro, se trata, como veremos, de una narración que surge en algún punto de la evolución humana gracias a la cultura; su génesis es externa, no interna, pues tomamos “lo que está afuera” (aprendemos), para urdir tal narración.<sup>21</sup>

---

<sup>21</sup> Varios autores, de distintas tradiciones, han abordado el problema del ego, poniendo a prueba la visión tradicional relacionada al cartesianismo, que considera que aquél es una sustancia pensante: Hume, Nietzsche, Kant, Sarte, Ricoeur, James, etc. No es mi propósito revisar tales argumentos en este trabajo.

Podemos adelantar que, en el caso del ego, se aplica lo mismo que apuntamos respecto de la conciencia: Dennett distingue entre conciencia humana y la de otros animales sin que esto implique una especie de dualismo; en todo caso hablaríamos de gradualismo<sup>22</sup>, como vimos al mencionar la evolución de la conciencia; esto es, en sus escritos tenemos que contextualizar, pues frecuentemente habla de ‘conciencia’ a secas, sin añadirle “humana” o “no-humana”. Pues bien, pasa lo mismo cuando se expresa en torno al ego. Pero tal cosa no implica que el mundo esté dividido en dos: ego humano, por un lado, y ego de los animales no humanos por el otro.

Veremos, pues, cómo el americano desarrolla su teoría del ego de manera paralela a su teoría de la conciencia; esto es, una parte se remitirá a las metáforas, un segundo paso a pruebas empíricas, intentando mostrar el poder explicativo del ego entendido como narración y, finalmente, veremos el aspecto evolutivo, también adaptacionista.

Podemos adelantar que algunas críticas sostienen que el ego que sale de las explicaciones dennettianas es demasiado parco, muy lejano al yo “cotidiano” (McCarthy, 2007: 247), sin explicar muy bien el aspecto social de la construcción de tal fenómeno, pero cuando hablemos de la evolución del ego veremos cómo la idea central de nuestro autor es introducir el gradualismo en el surgimiento del yo humano, no explicar con precisión cómo se construye socialmente.

---

<sup>22</sup> No queremos decir que dualismo y gradualismo sean incompatibles, pero sí nos parece importante hacer tal distinción ya que, comúnmente, llamamos dualismo a una postura cuya transformación (lo que aquí llamamos modelo del TC) implica presupuestos metafísicos distintos respecto de aquellos del gradualismo: uno, perteneciente al primero, es aquél que sostiene que la conciencia y el ego humanos son una especie de punto central, que, además, hace cualitativamente diferente al cerebro humano (de donde es plausible creer que surge la conciencia) respecto del cerebro de otras especies -tal es el dualismo que creemos incompatible con la postura filosófica de Dennett-, mientras que, el gradualismo dennettiano, por su parte, cree que puede rastrearse la continuidad entre los dos polos del ego de los que habla: el ego biológico y el ego cultural.

## **Las metáforas del ego**

Podemos comenzar con la metáfora que nuestro autor utiliza para pensar al ego humano, que es la del centro de gravedad, claramente relacionada con lo que ya hemos apuntado acerca de las insuficiencias explicativas del modelo del Teatro Cartesiano:

What is a self? I will try to answer this question by developing an analogy with something much simpler, something which is nowhere near as puzzling as a self, but has some properties in common with selves [...] What I have in mind is *the center of gravity* of an object [...] This is a well-behaved concept in Newtonian physics. (Dennett, 1992: 1).<sup>23</sup>

Los aspectos que nos interesan del concepto de centro de gravedad son:

1.- Un centro de gravedad no es un átomo, ni una partícula subatómica, carece de propiedades físicas excepto por ser localizable espacio-temporalmente; de hecho no es ningún objeto físico en el mundo. Su característica principal es que se trata de un objeto abstracto que, sin embargo, juega un papel bien definido en la física clásica, podemos decir que es una ficción teórica (*a theorist's fiction*). Si considerásemos al ego como un átomo en particular, estaríamos cometiendo un error categorial, nos dice el americano (Dennett, 1992: 6).

2.- Otra característica de los centros de gravedad es que podemos manipularlos, podemos cambiar el centro de gravedad de un objeto, de modo que aunque se trata de un objeto abstracto, *it has a spatio-temporal career, which I can affect by my actions. It has a history, but its history can include some rather strange episodes. Although it moves around in space and time, its motion can be discontinuous* (Dennett, 1992: 2).

---

<sup>23</sup> El escrito se encuentra en lenguaje HTML, paginación de Word fuente 12.

3.- El status científico del concepto de centro de gravedad no es desacreditado por que sea una ficción teórica, por el contrario, es una asombrosa ficción teórica, con un legítimo lugar en la física clásica.

Los tres puntos anteriores, hablando de los centros de gravedad, aplican al ego humano, que es producto de la conciencia: es una ficción teórica, es maleable y puede ser *localizado* espacio-temporalmente; no es un punto central, pero sí juega un papel importante en la economía cognitiva del animal humano o no humano al que se le atribuye.

El terreno en el que el ego humano se interpreta no es desde la física (aunque podría ser visto desde allí, como veremos al hablar de las tres actitudes o estrategias de Dennett, al llegar a la Actitud Intencional), es en este primer sentido –desde “el exterior”, podríamos decir– que el ego humano es un centro de gravedad *narrativa*:

A self is also an abstract object, a theorist's fiction. The theory is not particle physics but what we might call a branch of people-physics; it is more soberly known as a phenomenology or hermeneutics, or soul-science (*Geisteswissenschaft*). The physicist does an *interpretation*, if you like, of the chair and its behavior, and comes up with the theoretical abstraction of a center of gravity, which is then very useful in characterizing the behaviour of the chair in the future, under a wide variety of conditions. The hermeneuticist or phenomenologist –or anthropologist– sees some rather more complicated things moving about in the world –human beings and animals– and is faced with a similar problem of interpretation. It turns out to be theoretically perspicuous to organize the interpretation around a central abstraction: each person has a *self* (in addition to a center of gravity). (Dennett, 1992: 3).

Es así que el ego es atribuido; podemos ver esto, además, en la psicología: el terapeuta “extrae” y analiza la narración de su paciente, hallando contradicciones, supuestos no analizados, detalles olvidados, etc., y encarándolo con ellos.

Así, a modo de centro de gravedad, el ego humano funciona como un capitán (ficticio)<sup>24</sup> para el cuerpo donde se aloja, que a “simple vista” –perspectiva

---

<sup>24</sup> No se trata del “dogma del fantasma en la máquina” del que habló Gilbert Ryle, recordemos que para Dennett no hemos de pensar en homúnculos, sino en un pandemónium, como vimos en el

“exterior”– parece uniforme, constante, a la manera de un continuo espacio-temporal. Pero el ego humano es más poroso de lo que puede parecer a simple vista, como veremos.

El otro aspecto –“interior”–, es el que nos es íntimamente familiar, es la auto-hermenéutica, la auto-interpretación, la historia que nos contamos a nosotros mismos acerca de nosotros mismos.

In fact we have to posit selves for *ourselves* as well. The theoretical problem of self-interpretation is at least as difficult and important as the problem of other-interpretation (Dennett, 1992: 4).

Esta distinción entre las perspectivas “interna” y “externa” es mencionada simplemente para mayor claridad, no es ninguna distinción epistemológica, nos ayuda para lo mencionado por nuestro autor: cuando vemos a otro ser –humano o no– actuando de acuerdo a fines y medios, tendemos “naturalmente”, nos dirá Dennett, a atribuir un jefe interno. Es lo que hacemos todo el tiempo en nuestras relaciones con otros (humanos o no humanos).<sup>25</sup>

La perspectiva interna es la auto-atribución y reconocimiento de la propia narración, que ha llevado tradicionalmente a la excesiva confianza en la perspectiva de la primera persona, como vimos al hablar de la conciencia.

Notaremos mejor de lo que estamos hablando con ‘centro de gravedad narrativa’ si mencionamos el ejemplo de la ficción: si abrimos el cuento “la gallina degollada”, de Horacio Quiroga, y uno de los personajes dice: “llámame Mazzini”, ¿a quién le llamamos Mazzini?, ¿al escritor del cuento?, ¿al papel en el que se encuentra el personaje ficticio?, ciertamente no. Llamamos Mazzini al personaje, a

---

primer apartado. En el capítulo III hablaremos un poco más acerca de este capitán ficticio y de cómo para nuestro autor el ego es, fundamentalmente, autocontrol.

<sup>25</sup> La AI es una herramienta utilizada para intentar comprendernos, funciona bien en términos generales (como podemos observar cotidianamente atribuyendo creencias y deseos a otros) por motivos evolutivos, nos dice nuestro autor; es una caja negra, no sabemos si funciona por una especie de sistema de atribución de creencias o por rasgos, regiones o células cerebrales específicas (como las células espejo, por ejemplo).

la ficción, a una historia tejida por el autor, que, dicho sea de paso, mientras mejor sea el escritor mejor será la creación, más dimensiones psicológicas y de personalidad tendrá, más complejidad y riqueza hallaremos.

Pero la principal diferencia entre los centros de gravedad y los personajes de la ficción, es la que nos señala que un personaje de ficción es indeterminado en muchos aspectos, no tiene sentido preguntar si Mazzini usaba una playera roja o tenía ojos azules, si el escritor no lo especifica. El concepto de centro de gravedad, por el contrario, está ceñido a la teoría científica de la que es una parte. No pasa lo mismo con la narración, pues si el escritor hiciera un cuento posterior donde se hablara de tales detalles, es claro que no podríamos de hablar de indeterminación. En el mismo sentido, nosotros, como ego-narración, podemos ir modificando y añadiendo aspectos y detalles, aunque no podamos modificar lo ya “escrito”:

I propose that this imagined exercise [modifying our narrative path and therefore our ego] is actually a familiar exercise. That is the way we treat each other; that is the way we are. We cannot undo those parts of our pasts that are determinate, but our selves are constantly being made more determinate as we go along in response to the way the world impinges on us. Of course it is also possible for a person to engage in auto-hermeneutics, interpretation of one's self, and in particular to go back and think about one's past, and one's memories, and to rethink them and rewrite them. This process does change the "fictional" character, the character that you are [...] This would be an utterly mysterious and magical prospect (and hence something no one should take seriously) *if the self were anything but an abstractum* (Dennett, 1992: 8).

No podemos decir, además, que los personajes de ficción necesiten un yo “real” para existir, podemos imaginar, por ejemplo, una máquina que escriba cuentos (esto es algo que se ha intentado)<sup>26</sup>, y lo mismo aplicaría si esa máquina fuera programada para que esos cuentos, esa narración, se la atribuyera a sí misma. El ego humano, pues, no es el punto central, el proveedor de significados que valora, experimenta, agente y condición de posibilidad de los fenómenos. Ya hemos visto los problemas que presenta la idea del TC refiriéndonos a la conciencia; pues bien, el ego humano, de acuerdo con el profesor de Tufts, no es ningún

---

<sup>26</sup> En 1976, James Meehan, del laboratorio Schank, en Yale, creó un programa para escribir cuentos llamado TALESPIN.

homúnculo o punto central, sino un producto de la conciencia, una ficción que se crea desde la cultura cuyo soporte es el cerebro.

"Where is the self?" a materialist philosopher or neuroscientist might ask. It is a category mistake to start looking around for the self in the brain. Unlike centers of gravity, whose sole property is their spatio-temporal position, selves have a spatio-temporal position that is only grossly defined. Roughly speaking, in the normal case if there are three human beings sitting on a park bench, there are three selves there, all in a row and roughly equidistant from the fountain they face. (Dennett, 1992: 6).

Parece pues, que hay un jefe interno, pero se trata de una ficción, ¿no pasa algo similar con muchas de las colonias de insectos (como las colonias de termitas), que parecen seguir un plan diseñado por un jefe?

Una analogía más de Dennett al respecto es la de los Estados: por ejemplo, a los EE. UU., donde no tendría sentido buscar a un yo definido, y sin embargo podemos decir que tal país tiene recuerdos (como la guerra de Vietnam), o que odia al comunismo, etc.; lo que se encuentra, en todo caso, es un cúmulo de distintas agencias (o departamentos), operando con relativa independencia.

Así pues, hallamos en esa nación una suerte de cabeza de estado: un Presidente, que se supone representa a la nación al interior y al exterior, es el "vocero"; y aunque esto no significa que sin tal vocero el "cuerpo" deje de funcionar, podemos esperar que funcione mejor si está presente una figura tal. Un cuerpo humano, así, necesita –porque funciona mejor– una figura central o testaferro (*figurehead*), especialmente ante lo complejo de la vida social. Y antes de hallar tal figura, los candidatos suponen qué clase de persona se espera que sean o pretendan ser y se adaptan a tal expectativa. Como hay muchas versiones de lo que se espera, hay distintas versiones hechas de lo que se pretende.

Tenemos, así, que un ser humano crea primero uno o más egos y luego se elige uno. La influencia externa (padres, amigos, enemigos, etc.) tiene un papel clave en la configuración de lo que se supone significa ser uno mismo. Normalmente sólo uno de esos egos triunfa y persiste como narración del "ego real" (si por

distintas circunstancias, esos egos muestran un balance, tenemos los casos de Desorden de Identidad Disociada, que veremos enseguida).

Al decir, por ejemplo, *How can I tell what I think until I see what I say?* (1989b: 17)<sup>27</sup>, ese “I” se refiere a lo mismo, y Dennett sugiere que sólo se trata de una “tradicón gramatical”, que depende del hecho de que el pensamiento expresado en la oración se auto-confirma: *what “I” (my self) thinks is what “I” (my language apparatus) says.* (1989b: 17), sin embargo, no puede haber garantía de la existencia de ese ego solitario, ni de que un oyente o interlocutor admita un solo ego: supongamos, nos propone Dennett, que en diferentes momentos, diferentes sub-sistemas dentro del cerebro producen “racimos” de discursos que simplemente no pueden ser interpretados como productos de un solo ego.

La idea de “elección” de un vocero o *Head of Mind* no es, nos dice Dennett, implausible: eventos similares a elecciones ocurren todo el tiempo en el cerebro humano; por ejemplo, cuando el sistema visual recibe una imagen distinta en cada ojo, primero hay un intento de fusión de las imágenes y, ante la imposibilidad de hacer tal cosa, se da la “rivalidad binocular” (*binocular rivalry*), donde una de las imágenes es suprimida para dar paso a la otra, tal como lo explicamos más arriba con el fenómeno del metacontraste (*Cf. p. xv*). Así, al nivel de la neurofisiología visual, se tiene evidencia que apoya la preferencia humana por la unicidad por sobre la completud.

### ***Respecto del ego como narración***

Es probable que la última parte del apartado anterior, donde hablamos del ego como centro de gravedad narrativa, se aprecie con más claridad si nos referimos brevemente al Desorden de Personalidad Múltiple o Desorden de Identidad Disociada (DID), donde el paciente parece ser compartido por varios egos, cada

---

<sup>27</sup> El texto se encuentra en lenguaje HTML, paginación de Word fuente 12.

uno con su propio nombre y autobiografía. La complejidad y el grado de detalle en la construcción de las distintas personalidades puede llegar a niveles asombrosos (Dennett, 1991: 419). El principio “un ego por cuerpo” es la norma, pero en estas condiciones parece no aplicarse. Este fenómeno no reta, de acuerdo con Dennett, nuestras presuposiciones acerca de lo que es metafísicamente posible, sino que reta a aquellas presuposiciones acerca de lo humanamente posible, acerca de la crueldad humana, por un lado (pues estos casos se presentan normalmente bajo condiciones de abuso y de confusión extrema) y, por el otro, de la creatividad para ser capaz de construir distintas narraciones.

Ahora bien, si el ego humano fuera una región o módulo cerebral y pudiera hallarse en un punto específico del cerebro ¿cómo podría darse cuenta de esa partición o división de un elemento físico que se observa en los pacientes con DID? Este proceso creativo parece ser la obra de un supervisor, de un controlador central, pero, como ya lo mencionamos, pasa algo similar con muchas de las colonias de insectos. Para Dennett, como puede verse, el DID es mejor explicado si partimos de la hipótesis de que el ego humano es una narración.

Podemos encontrar otra pista acerca del status de narración del ego con los fenómenos analizados por Michael Gazzaniga (2008: 289) en el caso de pacientes epilépticos que se les ha realizado una comisurotomía (cerebro dividido), que consiste en la desconexión de los dos hemisferios cerebrales mediante la sección de las fibras nerviosas que interconectan a ambos hemisferios cerebrales (cuerpo calloso); con este procedimiento la frecuencia de los ataques epilépticos se reduce y se evita que ambos hemisferios simultáneamente se vean afectados durante una crisis.

Después de la comisurotomía, los pacientes muestran “dos cerebros”, por decirlo de algún modo: ambos hemisferios son prácticamente independientes, por lo que cada uno puede “informar” a otras regiones cerebrales cosas distintas en una misma situación. Los pacientes no muestran señales de división en cuanto a

rasgos psicológicos, pero sí motores: *While both hemispheres can generate spontaneous facial expressions, only the dominant left hemisphere can generate voluntary facial expressions.* (Gazzaniga, 2008: 291).

De acuerdo con Gazzaniga, la mente es una compleja combinación de sistemas parcialmente autónomos, y puede ser que un sistema o subsistema no tenga acceso a otros todo el tiempo, lo cual crea problemas de comunicación interna que son resueltos de diversas maneras. Ahora bien, *why don't split-brain patients have dual consciousness? Why aren't the two halves of the brain conflicting over which half is in charge? Is one half in charge? Is consciousness and the sense of self actually located in one half of the brain?* (Gazzaniga, 2008: 290).

Dennett cree que la respuesta tiene que ser evolutiva y, por lo tanto, especulativa: *sometimes talking and listening to yourself can have wonderful effects, not otherwise obtainable* (Dennett, 1984: 30-40). Recordemos el “cuento” ya mencionado para la evolución de la conciencia.

Gazzaniga se adhiere a una postura similar (aunque habla de módulos<sup>28</sup>):

Taking an evolutionary perspective [...] humans have evolved from creatures whose behavior was determined by modular systems with little integration. As these systems evolved and became more complex, consciousness emerged as a way of exerting executive control over diverse functions (Gazzaniga, 2000: 1355, 1362).<sup>29</sup>

Las investigaciones de Gazzaniga muestran la gran variedad de recursos exhibidos por el hemisferio derecho cuando se le encara a un problema de comunicación: en un experimento se pide a los pacientes que identifiquen por su forma a un objeto, dentro de una bolsa oscura, con la mano izquierda (que es

---

<sup>28</sup> *The neocortex in all species consists of an array of iterative neuronal groups called, interchangeably, radial columns or modules. These groups share a common intrinsic and extrinsic connectivity and subserve the same function [...] The larger the cortex in a given species, the larger the number of participating columnar units* (Gazzaniga, 2000: 11).

<sup>29</sup> Aunque Dennett no es un partidario de la modularidad como esquema rígido (entendida ésta como se puede ver en la nota 28), pues la misma función (o funciones cerebrales similares) pueden ser realizadas por distintas regiones, éste es uno de los puntos del Modelo de las Versiones Múltiples.

controlada principalmente por el hemisferio derecho) y luego que lo identifiquen verbalmente (el control y producción del lenguaje es hallado principalmente en el hemisferio izquierdo)<sup>30</sup>; en personas “normales” (sin “cerebro dividido”); esta tarea es aparentemente muy simple, pero los pacientes con comisurotomía saben de qué objeto se trata, pero no pueden verbalizarlo. Hay variantes con este experimento, siempre con resultados similares (Gazzaniga, 2000: 1358).

El punto es que la solución surge rápidamente: las sensaciones táctiles son representadas contralateralmente (van al hemisferio contrario), pero las señales de dolor son representadas ipsilateralmente (van a los dos hemisferios cerebrales); esto es, gracias a la configuración física del sistema nervioso, la táctica es provocar sensaciones dolorosas (si hubiera un objeto puntiagudo en la bolsa, por ejemplo) que llegan a ambos hemisferios, de modo que el hemisferio izquierdo “sabe” que hay algo puntiagudo en la bolsa, lo cual basta para que comience a adivinar, esto porque el hemisferio derecho emite señales controladas (de temperatura que aumenta en la mano derecha, por ejemplo); eventualmente, después de este extraño juego de adivinanzas entre los dos hemisferios, el “único” sujeto, –el habitante del cerebro–, dice el nombre del objeto en la bolsa.

Si buscamos al ego humano en el cerebro y/o suponemos que hay sólo un ego por cuerpo, difícilmente podremos explicar fenómenos como los que Gazzaniga nos narra en sus experimentos. Con la explicación de Dennett *when a human being's behavioral control system becomes seriously impaired, it can turn out that the best hermeneutical story we can tell about that individual says that there is more than one character "inhabiting" that body* (Dennett, 1992: 13). Parece que no hay, pues,

---

<sup>30</sup> Aunque las cosas no son tan simples, como nos hace notar Gazzaniga (no en todos los procesos cerebrales aparece la misma distinción, aunque sí se da en los que mencionamos), lo importante es mostrar los “trucos” o la multifuncionalidad de que se vale el cerebro, sin que haya un jefe interno: *You may have been taught in anatomy lectures that the right hemisphere of the brain controls the left half of the body and left hemisphere controls the right half of the body. Of course, things are not quite that simple. For instance both hemispheres can guide the facial and proximal muscles, such as the upper arms and legs, but the separate hemispheres have control over the distal muscles (those farthest from the center of the body), so that, for example, the left hemisphere controls the right hand* (Gazzaniga, 2008: 290).

un yo único, un núcleo pensante, sino una serie de trucos y vías en las que el cerebro responde a los estímulos del ambiente y crea, de un modo u otro, soluciones.

Dennett (1991), for example, has suggested that the right hemisphere (RH) has, at best, a “rudimentary self.” We, too, believe that the RH has more limited cognitive abilities than the left; but we feel it is crucial to accurately assess the upper extent of its cognitive function (Gazzaniga, 2000: 1355).

Nuestro ego, más bien, es *una historia* que nos define a nosotros mismos, una ficción teórica, *un punto luminoso de auto-representación* (Dennett, 1991: 429). La diferencia entre el *yo* y el *ello* es la utilización que se le da y no la *forma* en que aparece. El cerebro recoge y organiza información acerca de tal narración del mismo modo que recoge cualquier otra información, pues aquél es una especie de máquina producto de la evolución, como mencionamos al inicio de este trabajo.

Esto nos ayudaría a explicar, además, los cambios en la personalidad a través del tiempo: ¿qué nos hace pensar que somos la misma persona que cuando teníamos 10 años? La respuesta es: la persistencia de la narración, pues aunque algunas creencias se modifiquen o desaparezcan el núcleo narrativo se sostiene en la memoria y en las relaciones interpersonales, el tejido de creencias surgido, gracias al lenguaje principalmente, le da al ego humano la estabilidad sin requerir límites rígidos.

Como ya mencionamos en el caso de la metáfora del centro de gravedad, el ego entendido así es una trabajosa ficción teórica que juega un importante papel en la economía cognitiva, cuya continuidad depende de la coherencia o persistencia de la narración. Es, por ello, el modelo más importante del agente, base de los demás modelos, de ahí que muchas veces se le haya considerado origen y no producto.

## ***La evolución del ego***

Hemos mencionado ya que la visión imperante que tenemos respecto del ego humano es la heredada por Descartes: una especie de fantasma inmaterial al que pertenece el cuerpo y que lo controla; sin embargo, y con el surgimiento de un tipo de materialismo que afirma que la mente no es más que el cerebro, se ha rondado la idea de que el ego humano, al no ser tampoco más que el cerebro, ha de ser localizable en alguna región de éste; pero ¿es esta idea del ego una poderosa ficción, un mito que seguimos contándonos a nosotros mismos a pesar de los avances de la ciencia que lo desacreditan?

De acuerdo con la teoría del ego de Dennett, nos hallamos frente a un mito capacitador, una especie de hipótesis como lo han sido el éter, el alma o el flogisto; buscar al ego en alguna región cerebral sería poco fructífero (como parecen mostrar los experimentos de Libet que veremos, con el presupuesto de que hay un lugar donde surgen, además y de algún modo, la conciencia y la voluntad libre):

You enter the brain through the eye, march up the optic nerve, round and round in the cortex, looking behind every neuron, and then, before you know it, you emerge into the daylight on the spike of a motor nerve impulse, scratching your head and wondering where the self is. (Dennett, 1989b: 3).

Tenemos, pues, que comenzar por otro rumbo, pero ¿cuál? Dennett propone seguir un modelo evolutivo. Si hay tal cosa como criaturas con una cosa llamada ego, y tales cosas son algo, existen en algún sentido, hubo por lo tanto un tiempo –tal vez hace miles de millones de años–, donde los egos no existían; luego, ha de haber una historia que nos cuente cómo es que surgieron criaturas con ego, que emerjan de criaturas que no lo tienen a seres que sí lo tienen, pasando por un proceso o serie de procesos que permitieron tal cosa.

Es claro, sin embargo, que es imposible conocer tal historia a detalle, por lo que nos hallaremos frente a una historia especulativa que habrá de apoyarse en los

hechos disponibles y las analogías. Dennett reconoce las limitaciones de este enfoque, señalando que recurrirá a metáforas y simplificaciones. Esta historia será, por otro lado y de acuerdo con nuestro autor, similar a otras que la ciencia comienza a contar, como la del sexo:

[T]here is almost nothing sexy (in human terms) about the sex life of insects, oysters, and other simple forms of life, but we can recognize in their mechanical and apparently joyless routines of reproduction the foundations and principles of our much exciting world of sex. Similarly, there is nothing particularly *selfy* about the primitive precursors of human selves (Dennett, 1989b: 5).

El punto de partida es que la distinción entre egos surge de un principio fundamental: la biología comienza con la auto-preservación, con la emergencia de entidades que resistieron a la destrucción y al decaimiento, o sea, a la Segunda Ley de la Termodinámica, y que transmitieron a sus descendientes esa misma capacidad; dicho de otro modo, si la auto preservación es lo que se busca, los límites se vuelven importantes, pues no se querrá (ni se podría) preservar –en el sentido mencionado– al mundo entero. *You become, in a Word, selfish* (Dennett, 1989b: 6).

Esta es la forma primordial del egoísmo, una de las características de la vida. Es en esta distinción entre ego y “resto del mundo” donde se encuentra el origen del ego humano –y de todos los egos conocidos–, la barrera entre “dentro” y “fuera”, entre “mundo externo” y yo, es la que, según Dennett, se encuentra en el corazón de todos los procesos biológicos (se considera, por ejemplo, el caso de los anticuerpos como defensas contra “intrusos”).

Esta distinción fundamental, entre yo y mundo, tiene ecos incluso en nuestra psicología: tragar la propia saliva mientras se habla, por ejemplo, es distinto a que si uno escupiera en un vaso y luego esa saliva se introdujera de nuevo en la boca; el primer caso es normal, el segundo, repulsivo. Ejemplos como este, del psicólogo Paul Rozin, sugieren a Dennett (1991: 414) que hay una poderosa y

ciega resistencia –que parece surgir de la división mencionada– a ciertos actos que analizados con más detenimiento, no parecen tan peligrosos.

Apenas hace falta decir que estos ejemplos de Dennett no son muy fuertes, pues se refieren al aspecto psicológico o, por decirlo de otro modo, al ego humano (difícilmente concebiremos a una langosta analizando sus acciones), cuando de lo que se trata, precisamente, es de hallar ese vínculo, proceso gradual (evolutivo) o, en todo caso, salto, del ego biológico al ego humano –cultural– y no proponerlo como análogo desde el inicio, como me parece que hace nuestro autor.

Pero podemos quedarnos con la distinción básica: tenemos, pues, un ego primitivo, un ego “mínimo”, el biológico, que se muestra, por ejemplo, en la organización de los sistemas internos de los organismos: *The earliest control systems were really just body protectors* (Dennett, 1996: 64). En la misma línea, *an autonomic nervous system is not a mind at all but rather a control system, more along the lines of the nutritive soul of a plant, that preserves the basic integrity of the living system* (Dennett, 1996: 25).

Este ego “mínimo” no es una cosa dentro de un organismo, ni el organismo entero, sino una *abstracción* que señala que existe una organización tal que tiende a *distinguir, controlar y preservar porciones del mundo*, que mantiene límites. Una definición para este ego biológico sería la siguiente: tú eres aquello que controlas y proteges.

Esta idea del ego biológico puede parecer un tanto excesiva: ¿tienen entonces las amibas un ego?, ¿podemos hablar del ego de una bacteria? ¿Pero no es, como señala Dennett, un caduco antropomorfismo reservar ‘conciencia’ para los humanos? (Dennett, 1991: 448), del mismo modo, el concepto de ego nos puede servir para desarrollar nuestro punto: cuando hablamos de ego en bacterias, la idea fundamental es la misma que se aplica a los humanos, es decir, una división básica entre el ser biológico y el resto del mundo, fronteras físicas y químicas, y

aunque la idea de ego se ha referido normalmente a la conciencia refleja, hallamos una segunda distinción anterior a ella: el yo como *autocontrol* que se da en grados, como límites, como continuo espacio-temporal, como veremos al mencionar los experimentos de Libet.

De modo que, al hablar de ego biológico, de lo que se trata es de atacar lo atractivo que pueda tener una postura apoyada por *the conviction that no positive naturalistic account can be given of what we hold dear when we contemplate "the dear self"* (Dennett, 1984: 80).

Hasta aquí, sin embargo, podemos ver que este ego mínimo no es el que nosotros buscamos, no es esa suerte de fantasma que controla al cuerpo del que Descartes nos habló, no es el ego que experimentamos los humanos todos los días de nuestra vida consciente; ese ego biológico suena, aunque familiar, un tanto lejano de lo que tenemos en mente cuando decimos, por ejemplo, yo pienso, yo creo, yo soy...

Pero, como se puede ver, la cuestión es si podemos reconstruir la línea o ramificación de evolución o desarrollo entre esos aparentes dos polos del ego, el biológico (presuntamente más simple) y el humano (narración).

Para empezar, notamos que la diferencia más importante entre el ego biológico y el ego como narración reside en la vieja atribución del lenguaje al ego: *ants and hermit crabs don't talk* (Dennett, 1989b: 10). El mundo humano, ciertamente, está plagado de palabras, de lenguaje, y aunque puede decirse que la organización del cangrejo ermitaño, por ejemplo, implica una concha externa, no se auto-representan como animales con concha encontrada. Los humanos, por el contrario, estamos constantemente comprometidos en presentarnos ante otros y aun ante nosotros mismos, o más bien a *representarnos* ante otros y ante nosotros mismos por medio del lenguaje y hasta de los gestos. El entorno o medio

ambiente humano no contiene sólo comida y refugio, enemigos o congéneres, sino muchas palabras:

These words are potent elements [...] that we incorporate, ingesting and extruding them, weaving them like spider webs into self-protective strings of narrative. Our fundamental tactic of self-protection, self-control and self-definition is not building dams or spinning webs, but telling stories –and more particularly concocting and controlling the story we tell others –and ourselves- about who we are. (Dennett, 1989b:11).

La semejanza más importante entre otras especies y nosotros, por otro lado, es que así como las arañas no tienen que pensar consciente y deliberadamente en cómo tejer su telaraña, o algunas aves macho en cómo hacen nidos llamativos para atraer a las hembras, nosotros, los humanos (a diferencia de los narradores profesionales) tampoco imaginamos consciente y deliberadamente qué y cómo narrar, sólo lo hacemos; más aun, nuestra conciencia y ego son su producto, no su origen. Tal es el punto clave.

Hay que decir que, cuando nuestro autor habla de “narradores profesionales” (Dennett, 1989b. 11), se puede prestar a confusión: es la cultura (los memes) la que configura la narración, no hay ningún ego en el sentido de “narrador”, su fuente no es interna, sino externa. Los narradores profesionales narran sus creaciones, pero ellos mismos (su ego) está ya mediado culturalmente cuando narran y crean. Esto nos llevaría al problema de la agencia, preguntándonos en qué sentido podemos imputar responsabilidad a las personas y, en último término, a indagar quién o qué toma las decisiones –veremos esto *brevemente* al hablar de los experimentos de Libet, en el siguiente apartado– en ese sentido, parece que la idea de centro de gravedad narrativa suplanta a la idea del ego cartesiano como herramienta heurística.

Tanto el ego biológico como el humano son, pues, principios de organización, pero sólo el primero es origen, mientras que el segundo es producto. El camino que lleva de uno hacia el otro se traza, así, entre los grados de ego como (auto) control y (auto) definición, además del breve cuento adaptacionista (Just So Story) que ya

mencionamos al hablar de la evolución de la conciencia.

Los flujos de narración de los que se compone el ego humano, al surgir de un mismo cuerpo, producen el *efecto* de que los oyentes o interlocutores postulen una ficción útil: un centro de gravedad narrativa, un notable, robusto y casi tangible *atractor de propiedades*. Este ego surge, reiteramos, desde un tejido de palabras y de actos que se hace sin realmente saberlo. El ego humano es la más extraña construcción del mundo animal, pues cada individuo de esta especie construye un yo, tejiendo una red de palabras y de hechos (Dennett, 1991: 416).

Como ya hemos mencionado, para Dennett el punto clave de los egos humanos es que están compuestos de memes, de cultura, particularmente de lenguaje, de ahí que la naturalización del ego que Dennett propone nos parezca muy lejana, tal vez un tanto parca, pero ésa es precisamente la idea del ego biológico como punto de partida, y nos explica por qué el americano no abunda en la construcción social del ego.

The idea that a full-fledged self or person must be the outcome of processes of social interaction has been ably and persuasively explored by others, and I will not take the time to review their reasoning. Instead I will look at this process of the acquisition of selfhood from the simplified perspective of strategies of control (Dennett, 1984: 82).

Ahora bien, hasta aquí hemos visto cómo nuestro autor recurre a metáforas, analogías, pruebas científicas y análisis filosófico de algunos presupuestos metafísicos, y tratamos de mostrar cómo el aspecto importante de los argumentos de Dennett es el evolucionista, pues las metáforas, las analogías y los ejemplos de la ciencia se modifican y ciñen (incluidos, claro, los “cuentos” adaptacionistas) a lo que considero la idea básica que sostiene el profesor de Tufts: llevar a cabo una inversión del razonamiento a la manera del evolucionismo –donde la mente aparece al final– en aspectos tan importantes como el ego y la conciencia, donde estos fenómenos, tradicionalmente entendidos como origen (de las cualidades fenoménicas, de la agencia, del ordenamiento de las percepciones, de las facultades valorativas y, por lo tanto, de la moral, etc.), son mejor pensados,

explicados y clasificados como producto, como el punto más reciente de una larga cadena evolutiva que ha llevado a lo que hoy somos como especie.

Tal vez nunca lleguemos a saber “realmente” qué es el ego, pero es razonable analizar lo que tenemos hasta ahora, para qué nos ha servido, para qué lo queremos y si nos sigue siendo útil; de otro modo, hemos de modificar viejos presupuestos y aventurarnos a lanzar hipótesis y teorías con mayor poder explicativo, como me parece que hace el americano.

Mencionamos al inicio de este trabajo que desde la perspectiva del modelo del TC, no hay una distinción clara entre conciencia y ego, parece que ambos se dan en el mismo instante, ninguno es condición de posibilidad del otro: el ego es autoconciencia y prueba de la propia existencia. En Dennett, por el contrario, el ego humano es producto de la conciencia, una narración de origen biológico y podemos distinguir sin mayor problema: el cerebro crea conciencia y, configurada por la cultura, ésta produce un aparente jefe interno, sin que esto implique una división fundamental con el resto de las especies, pues muchas de ellas pudieren tener conciencia y ego, aunque en distinto grado.

[L]anguageless animals "are not conscious at all" (as Descartes insisted)? [...] Notice that [this question] presupposes something we have worked hard to escape: the assumption that consciousness is a special all-or-nothing property that sunders the universe into two vastly different categories: the things that have it [...] and the things that lack it. The theory of consciousness we have sketched allows for many variations of functional architecture, and [...] the presence of language marks a particularly dramatic *increase* in imaginative range, versatility, and self-control.

While a bat, like even the lowly lobster, has a biological self, it has no selfy self to speak of — no Center of Narrative Gravity, or at most a negligible one. No words-on-the-tip-of-its-tongue, [...] no reflections on what it is like to be a bat; one does not do a favor to nonhuman animals by trying to obscure the available facts about the limitations of their minds (Dennett, 1991: 447-448).

Pero el darwinismo ortodoxo de Dennett resulta ser un arma de dos filos: si bien la idea de naturalizar a la conciencia y al ego me parece una ruta o programa de investigación plausible, el “salto” que el autor de *Consciousness Explained* da

hacia una especie de darwinismo universal adhiriéndose a la teoría de los memes, propuesta originalmente por Richard Dawkins, parece ser no sólo poco fructífero, sino materia de una serie de confusiones terminológicas y de hipótesis sin confirmar, que parecen poco estimables y que se asemejarían a aquellos discursos que Locke, citado por Dennett, tanto criticaba.

Al respecto, veremos un ejemplo del análisis de los presupuestos metafísicos implicados en los fenómenos de ego y de conciencia, apoyado en los experimentos de Libet, ahondando en el aspecto del yo como autocontrol. Concluiremos esta sección analizando cómo entiende Dennett el adaptacionismo y la relación que éste guarda con la Actitud Intencional, para finalmente abordar la teoría memética.

## Capítulo III

### **Los experimentos de Libet**

En 1985, los experimentos acerca de la presunta relación entre el cerebro y la conciencia que realizó el neurólogo norteamericano Benjamin Libet causaron impacto no sólo entre los científicos de su campo, sino entre psicólogos, sociólogos, filósofos, etc., por lo extraño de sus resultados. La consecuencia más notable de los famosos experimentos de Libet fue que parecían mostrar que la voluntad libre (*free will*) no existe sino como una ilusión o un autoengaño humanos. El punto es que tales experimentos se volvieron un caos de interpretaciones a favor, en contra, exigencias de métodos más precisos, análisis de la situación experimental, intentos de refutación, extensión de investigaciones, etc.

Creemos que estos experimentos muestran los presupuestos metafísicos en las neurociencias, punto que habla de la relación entre ciencia y filosofía que, de acuerdo con nuestro autor, sería la más fructífera<sup>31</sup> (recordemos también, en este mismo sentido, al fenómeno *phi*, al fenómeno del metacontraste, los experimentos con el cerebro dividido y al Desorden de Identidad Disociada o DID). Esto es, el modelo del TC está lejos de ser una postura olvidada: persiste no sólo en el sentido común, sino también en contextos científicos.

---

<sup>31</sup> Al respecto, Dennett señala: *I like [...] the distinction between philosophy in cognitive science and philosophy of cognitive science. In principle, you can do one without the other, and some work in the field can be quite clearly given just one of these labels, but in general I think that good philosophical contributions to cognitive science must work on both projects at once: it is a mistake to think, for instance, that philosophy of cognitive science is “just for philosophers” or “just for the lay public” a collection of antidotes to faulty inferences about the import of contemporary scientific work. [...] Philosophers often comically misjudge their competence to evaluate concepts, arguments, theories with which they have only the most passing acquaintance, and they are prone to coming up with what might be called bumblebee deductions -proving from “first principles” that bumblebees cannot possibly fly- [...] I think we philosophers can make [...] contributions; at the very least we can sharpen and clarify the alternatives, rendering the suspect claims clear enough to be refuted, which is definitely progress.* (Dennett, 2009: 231).

Los experimentos de Libet los encontramos claramente narrados en el siguiente artículo de la revista *Acta Philosophica*<sup>32</sup>:

El potencial de preparación o disposición (PD o *readiness potential*) es un cambio eléctrico en determinadas áreas cerebrales que precede a la ejecución de una acción futura. Libet y sus colaboradores diseñaron un experimento dirigido a conocer la relación temporal que existía entre el PD, la conciencia de la decisión de actuar y la ejecución del movimiento. La pregunta a la que intentaban responder era la siguiente: ¿cuándo aparece el deseo consciente o la intención (de llevar a cabo una acción)?

Para responder a la pregunta había que medir en qué momento preciso se generaban los PD, la conciencia del deseo de hacer un movimiento y la activación de los músculos efectores, de modo que permitiera secuenciarlos. [...] Los sujetos debían llevar a cabo un movimiento de la articulación de la muñeca en el momento en que sintieran el impulso de hacerlo. Para resolver el problema de la medición, los investigadores utilizaron un reloj, mucho más rápido que los normales, en el que un punto recorría la esfera entera en 2,6 segundos [...] El sujeto podía señalar cuál era la posición del punto en el preciso momento en que era consciente del deseo de mover la articulación examinada.

Para comprobar la precisión del informe de los sujetos, llevaron a cabo unas pruebas. En ellas, el individuo se encontraba relajado y no realizaba ningún movimiento voluntario, sino que recibía un débil estímulo en la piel de una de sus manos. El estímulo era repetido al azar en diversas ocasiones y el sujeto debía señalar en qué momento, según el reloj diseñado, era consciente de aquél. La diferencia con respecto al tiempo real, que los observadores conocían, tenía un error tan sólo de unos 50 milisegundos.

La conciencia del deseo de mover la muñeca se da 150 milisegundos antes de la activación del músculo. Por lo tanto, resultaba claro que los procesos cerebrales que preparaban el movimiento voluntario comenzaban 400 milisegundos antes de la voluntad consciente de actuar.

Libet y colaboradores concluyeron que “la secuencia del proceso volitivo ‘para actuar ahora’ puede aplicarse a todos los actos volitivos, independientemente de su espontaneidad o de la historia anterior de deliberaciones conscientes”. (Murillo; Jiménez-Amaya, 2008).

Como fue mencionado antes (*Cf. p. xliv.*), para Dennett, quien tanto se interesará en los trabajos de Libet, la idea de un ego monolítico, centro de control de todos los procesos cerebrales, un agente racional y unificado, es errónea: se trata del supuesto de que hay un lugar central en el cerebro donde se toman las decisiones conscientes, la *res cogitans*, con el ego como dueño del cuerpo y de sus percepciones. Por otro lado, y a pesar del materialismo dominante en las

---

<sup>32</sup> He decidido presentar los experimentos con este artículo, porque además de claro, nos ahorra los detalles técnicos sin dejar de lado los puntos importantes.

neurociencias, esa fuerza, que parece obligar a buscar un punto central en la conciencia o una línea divisoria entre lo consciente y lo no-consciente, parece seguir vigente. Es lo que el americano llama la *falacia del Teatro Cartesiano*: suponer que a menos que haya un punto central, no existen la conciencia, ni el ego.

### ***Medición y arribo de los contenidos a la “conciencia”***

Así, la idea del punto central de la conciencia ha de abandonarse pues –entre otras cosas– los avances en las neurociencias muestran cada vez con mayor fuerza que sostener un punto central para la conciencia es insostenible y, aunque frecuentemente recurre a metáforas, aquí podremos ver una parte de la teoría dennettiana del ego aplicada a un problema práctico científico (la parte negativa, acerca de cómo éste no es ningún punto central). Tal vez hablar de la distinción filo de y filo en buscar texto Se aplica, en particular, a la interpretación de uno que se dirige directamente a los presupuestos metafísicos, tanto del científico en cuestión cuanto de muchos de los intérpretes de los experimentos, a condición de que aceptemos que, hasta ahora, no se ha hallado algo así como un “núcleo jefe”.

Las respuestas, interpretaciones y objeciones han sido, como ya mencionamos, muchas y muy variadas; van desde la situación experimental hasta el análisis conceptual de ‘conciencia’, de ‘voluntad libre’, de ‘percepción’, etc. Pero podemos englobar los tipos de respuesta en un grupo común: los que centraban su atención en el procedimiento técnico, pues lo más normal fue –y ha sido– suponer que los tiempos de medición en los experimentos habían sido erróneos. De acuerdo con esta postura, alguno de los sistemas de medición en el recorrido de los contenidos de la percepción entre los músculos del brazo, el Potencial de Disposición, los ojos y la conciencia, había fallado.

Sin embargo, a pesar de su éxito, sobretodo mediático, este tipo de experimentos no han sido abandonados: en abril de 2008 se realizaron experimentos similares (poniendo aún más énfasis en asegurar la precisión de la situación experimental) por parte de John-Dylan Haynes del Instituto Max Planck, en Leipzig, arrojando los mismos resultados (Soler Gil, 2009).

La interpretación de Libet es que la voluntad libre se reduce a una especie de “derecho de veto”, pues los sujetos experimentales se dan cuenta de sus propias intenciones unos cuantos milisegundos antes de que la acción se realice efectivamente:

I have taken an experimental approach to this question. Freely voluntary acts are preceded by a specific electrical change in the brain (the ‘readiness potential’, RP) that begins 550 ms before the act. Human subjects became aware of intention to act 350–400 ms after RP starts, but 200 ms. before the motor act. *The volitional process is therefore initiated unconsciously. But the conscious function could still control the outcome; it can veto the act.* Free will is therefore not excluded. These findings put constraints on views of how free will may operate; it would not initiate a voluntary act but it could control performance of the act. The findings also affect views of guilt and responsibility. (Libet, 1999: 47).

Pues bien, el punto de partida de Dennett para analizar la interpretación de los experimentos de Libet es una línea específica del libro *The Intentional Stance*, que aquél considera la más importante del mismo: *If you make yourself really small, you can externalize virtually everything* (Blackmore, 2005: 90). Esto es, en los presupuestos de los que partimos encontramos las implicaciones, por decirlo de algún modo (si la conciencia y el yo son un punto central, otras regiones cerebrales serán siempre periféricas, añadidos externos).

Los experimentos de Libet, pues, dependen de situar la intersección de dos trayectorias: la percepción visual del reloj y el surgimiento “en la conciencia” de estar doblando la muñeca. Tomando estas dos fuentes de datos –así como el subsecuente reporte del sujeto y los escaneos cerebrales–, y añadiendo el momento en el que se registra el PD, Libet concluye lo que ya hemos explicado. Pero Dennett nos invita a considerar tres posibilidades más de interpretación, que

corren sobre la línea que recién mencionamos, respecto del tiempo de medición entre la mano, la conciencia, los ojos y el PD:

a) Tú estás ocupado tomando tu libre decisión en la *facultad del razonamiento práctico* (donde se toman todas las decisiones) y tienes que esperar allí los contenidos visuales enviados por *el centro de la visión*. ¿Cuánto tiempo toma eso? Si la presión sobre el tiempo no es crítica, tal vez el contenido visual es enviado lentamente y está seriamente desfasado con respecto a su tiempo de llegada a la facultad del razonamiento práctico.

b) Estás ocupado viendo el reloj en el *centro de la visión*, y tienes que esperar a que la facultad de razonamiento práctico te envíe los resultados de su más reciente decisión. ¿Cuánto tiempo toma ese proceso? Esta puede ser otra transmisión demorada.

c) Estás sentado donde siempre: en el cuartel general (conocido también como el Teatro Cartesiano) y tienes que esperar a que los informes tanto del *centro de la visión* como de la *facultad del razonamiento práctico* lleguen a donde todo se junta y surge la conciencia. Si alguno de estos puestos está más lejos que el otro, o transmite con más lentitud, estarás sujeto a ilusiones de simultaneidad.

De acuerdo con Dennett, estas posibilidades ayudan a mostrar los problemas de la postura de Libet, pues *the governing idea is that you can only act where you are, so if you are not in the faculty of practical reasoning when a decision is made there, you didn't make it. At best you delegated it* (Dennett, 2003: 42). Respecto de esta presuposición, tenemos dos alternativas: el *strolling you* y el *Out-of-touch-You*. En la primera opción, suponemos que tú decidiste conscientemente el movimiento de la muñeca, pero te tomó algo de tiempo (algunos milisegundos) llegar al centro de visión y observar qué posición tenía el punto indicador del reloj, o viceversa (de a y c).

La segunda explicación (c) es la que Dennett llama *Out-of-touch-You*, donde permaneces en el comando o cuartel general y esperas a que tus “funcionarios” envíen la información (confías en las redes físicas de información), de modo que tienes *control limitado* sobre tales fuentes de información; la red es una especie de circuito cerrado en el cual “tú” envías órdenes y esperas respuesta.

El punto es que los canales de transmisión en los que la conciencia confía (pues no tiene otra opción) pueden informar con desfase o incluso erróneamente: no podemos saberlo desde el “cuartel general” (recordemos nuevamente los fenómenos phi y del metacontraste ya analizados más arriba [Cf. p. xv]). Y las explicaciones que dan los sujetos experimentales de Libet toman como punto focal el momento en el que la orden de mover la muñeca “se hace consciente”.

Pero no tenemos por qué medir la conciencia (entendida como línea de meta) como el informe primordial emitido por el ego, pues múltiples regiones cerebrales están implicadas en las acciones realizadas por el cerebro; recordemos que ego y conciencia son productos del cerebro (una especie de máquina de miles de millones de años de evolución) y los errores se presentan, parece, cuando partimos de considerar al ego y a la conciencia como origen (organizador de las sensaciones, las percepciones, etc.)

Con explicaciones basadas en a, b y c se buscaría “salvar” a la voluntad libre: indagando en los desfases temporales y los tiempos de recorrido de las conexiones “periféricas” respecto de la conciencia. Y aunque, claro, puede haber más explicaciones que “salven” o no a la voluntad libre, Dennett nos pregunta:

Can't we just dismiss the whole sorry lot of them, on the grounds that these hypotheses are wildly unrealistic oversimplifications of what is known about how decision-making works in the brain? Yes indeed we could, and we should. But when we do that, we don't just dismiss all these fanciful hypotheses that could save “free will” in the face of Libet's data; we must also dismiss Libet's own hypothesis [and the others that] purport to show we only have “free won't”, [because they depend] on taking seriously the idea that you are restricted only to the materials you can get access to from a particular subregion of the brain (Dennett, 2003: 43).

Esto es, el problema con la idea de Libet, e intérpretes en general, es que presupone que la conciencia *no* tiene su origen en procesos “inconscientes”, y más aun, presupone que el cerebro (“inconsciente”) es capaz de –y es lo suficientemente listo– como para planear y dirigir, por ejemplo, los movimientos de la muñeca, pero *no* para decidir si tal acción se realiza o no (lo primero es “automático”, mientras que lo segundo es lo “propiamente humano”, ¿por qué situar a la voluntad libre en “donde surge la conciencia” y no en los múltiples procesos que tienen a la “decisión consciente” como a uno de sus últimos puntos?). Libet, notando esto, aunque sin abandonar su postura, añade:

The possibility is not excluded that factors, on which the decision to veto (control) is based, do develop by unconscious processes that precede the veto. However, the conscious decision to veto could still be made without direct specification for that decision by the preceding unconscious processes. (Libet, 1999: 53).

Pero, si la posibilidad de que la “conciencia” tenga origen en procesos “inconscientes”, el lapso o laguna de 350-400 milisegundos que ésta “espera” para encenderse y dar paso a la “ficción de voluntad libre” *no ha sido demostrada*. Esa aparente laguna sería, precisamente, parte de la preparación de la conciencia para tomar una decisión. Sabemos, por lo demás, que en circunstancias normales –nos dice Dennett– el cerebro comienza su trabajo evaluativo y discriminativo en cuanto recibe estímulos y trabaja en más de un proyecto a la vez, como intentamos mostrar al hablar de los modelos de la conciencia en Dennett, *v. gr.* el pandemónium y el Modelo de las Versiones Múltiples, sin que todo contenido tenga que pasar por el “foco” de la conciencia: ya lo hemos afirmado, la conciencia no es un “todo o nada”. De modo que la ventana de veto de 100 milisegundos de la que Libet nos habla es, nos dice Dennett, sólo *an artifact of the Cartesian Theater. When you distribute the work in time and space, you distribute the responsibility in time and space.* (Dennett, 2003: 44).

Esta ilusión del Teatro en la que muchas veces queda atascada la discusión anterior es vista con más claridad –cree Dennett–, si consideramos a la conciencia desde una perspectiva evolucionista: en la mayoría de las especies conocidas

hasta ahora, la “causación mental” nunca ha sido necesaria y, por lo tanto, no habría evolucionado ninguna capacidad de auto-monitoreo. En general, las causas trabajan bien en la obscuridad, nos dice el americano. La capacidad de los “outputs” de las especies no necesitan –o no han necesitado– ser experimentadas por nada ni nadie en absoluto. Estamos familiarizados con acciones apenas conscientes (automatismos), como cuando conducimos un auto pensando en qué vamos a hacer llegando a casa. Pues bien, un sinnúmero de situaciones y respuestas de gran sofisticación pueden residir en el sistema nervioso de una criatura y servir a sus necesidades sin que haga falta ninguna supervisión ulterior.

Un animal bien puede poner el alimento en su boca, golpear a su presa, etc., sin que haya un punto focal analizando la situación, sino una acción similar a la que realiza el organismo cuando regula la respiración durante el ejercicio (esta puede ser la verdad detrás de la suposición de que algunos animales son como robots o zombis).

Algunos otros animales desarrollan conductas de preparación de su ambiente que lo hagan más sencillo y seguro: nidos, escondites, túneles, etc. En la misma línea, cuando surge la necesidad, algunas criaturas desarrollan “trucos” para modificar y ajustar su ambiente más íntimo: sus propios cerebros, creando caminos y señales para uso posterior, controlándose y definiéndose. Es en este sentido que Dennett ha señalado que los humanos son animales que crean futuro y ficciones, cosa que no es menor en un mundo que sería más peligroso si la especie siguiera solamente el camino de prueba y error. ¿Y qué dio pie a tan importante salto?; ya hemos mencionado antes la respuesta de acuerdo con el profesor de Tufts: la comunicación, el lenguaje, la cultura.

### ***Breves consideraciones sobre la voluntad libre y el autocontrol***

El camino del autocontrol y la autodefinición es el que ha llevado a los humanos a crear el ego que observamos, una útil ficción que *da la impresión* de ser un centro de operaciones:

We wouldn't exist as Selves "inhabiting complicated machinery" [...] if weren't for the evolution of social interactions requiring each human animal to create within itself a subsystem designed for interacting with others. Once created, it could also interact with itself at different times [...] as Wegner vividly puts it, people became what they think they are, or what they find that others think they are, in a process of negotiation that snowballs constantly (Dennett, 2003: 47).

Con el lenguaje, comenzaron a existir los tipos de mentes que pueden transformarse a sí mismas, siguiendo nuevas directrices, comenzando nuevos proyectos, intentando nuevas vías de respuesta para un mismo problema. Eso es –nos dice Dennett–, precisamente, lo que significa que algo sea una mente, en contraste con un cerebro, pues aquélla es el sistema de control.

Es cierto que los animales no-humanos también pueden decidir, por ejemplo, un ave puede planear su vuelo hacia arriba o hacia abajo, en un sentido o en otro, con esta o aquella velocidad, pero a los animales humanos, en contraste, además de que se nos pueden pedir cosas de manera muy precisa, detallada y compleja por medio del lenguaje, somos capaces de instruirnos a nosotros mismos del mismo modo, de pedirnos y darnos razones para esto o para aquello, porque nuestro ego ha potenciado precisamente esas capacidades, tanto las de comprensión de los demás, cuanto de nosotros mismos (aunque todo ello sólo sea un ideal). En todo caso, queda claro que, desde una perspectiva evolucionista, hablamos de grados de control y definición.

La voluntad libre, pues, no es una capa anatómica del cerebro, sino una capa funcional, compuesta de algún modo en los micro-detalles del cerebro, pues podemos pedir a los otros y a nosotros mismos hacer cosas de manera precisa y controlada. Es esto lo que nos hace distintos como especie, pues podemos tomar

decisiones que afectan nuestra vida y la de otros teniendo conocimiento de tales circunstancias; esto es, se nos pueden pedir razones para nuestros actos *en un sentido moral*.

Es, en este sentido, que aparece el agente humano: podemos pedirle –o exigirle– razones para sus acciones, y podemos también ofrecer argumentos, precisamente porque es sólo socialmente donde el ego cobra su dimensión reflexiva (algo que no sucede por ejemplo con los ferales o niños salvajes) y desde allí se puede analizar e imputar responsabilidad.

Y, si por circunstancias particulares, decimos “no era yo el que realizó esa acción”, es una mala salida: difícilmente se trataría de una disrupción de autocontrol como la que observamos, por ejemplo, en pacientes esquizofrénicos. Si consideramos solamente a ese ego aislado, en nuestros esfuerzos por ponerlo bajo el microscopio, de las pruebas experimentales de Libet, entonces esa conciencia no es responsable, pues es una especie de máquina desconectada de sus actos pasados.

Precisamente esa es la función de lo que Dennett llama la ilusión del usuario (Dennett, 1991: 216): suponer y hasta cierto punto ejercer control sobre “apariencias” de mecanismos desconocidos, como los del cerebro, para proveernos de los medios necesarios para interconectarse con estados anteriores del ego.

Hemos llegado, así, a una parte importante de las consideraciones de nuestro autor respecto del ego humano: el autocontrol que surge a partir del manejo del lenguaje y la cultura humanas. Pues bien, este autocontrol, como hemos podido ver en este capítulo, viene también en grados, pues tratamos con autodefinición, podemos hablar de autocontrol reflexivo en humanos, pero también de autocontrol y autodefinición en otras especies cercanas, como los chimpancés, que poseen lenguaje y cultura (o protolenguaje y protocultura, un tema que no trataremos

aquí), además de aprendizaje social; pero esto es más difícil cuando hablamos de ostras o de bacterias. Añadimos, pues, un rasgo más del ego humano: la capacidad de autocontrol.

The perspectival trick we need in order to escape the clutches of the Cartesian Theatre is coming to see that I, the larger, temporally and spatially extended self, can control, to some degree, what goes on inside of the simplification barrier, where the decision-making happens (Dennett, 2003: 49-50).

Los primeros sistemas de control eran sólo protectores del organismo, límites, que fueron aumentando en complejidad:

In animals, [the] complex system of biochemical packets of control information was eventually supplemented by a swifter system, running in a different medium: traveling pulses of electrical activity in nerve fibers. This opened up a space of opportunities for swifter reactions, but also permitted the control to be differently distributed, because of the different geometries of connection possible in this new system, the autonomic nervous system. The concerns of the new system were still internal –or, at any rate, immediate in both space and time: Should the body shiver now, or should it sweat? (Dennett, 1996: 66).

[...] viewing the brain (and hence the mind) as one organ among many, a relatively recent usurper of control, whose functions cannot properly be understood until we see it not as the boss but as just one more somewhat fractious servant, working to further the interests of the body that shelters and fuels it and gives its activities meaning (Dennett, 1996: 76).

Y, para nuestra especie, añade Dennett:

[The] fundamental tactic of self-protection, self-control, and self-definition is not spinning webs or building dams, but telling stories, and more particularly concocting the story we tell others –and ourselves– about who we are (Dennett, 1991: 418).

Así, con el ego biológico hablamos de control y definición del organismo en contraposición con el resto del mundo, que va aumentando en grados (algunos pueden parecer saltos), hasta llegar a los tipos de mentes que se autocontrolan y se autodefinen en un sentido distinto por medio de la cultura. Desde allí es configurado el ego humano. La brecha entre nuestra especie y otras parece

bastante amplia, pero la diferencia es de grado y no de tipo, como hemos señalado varias veces.<sup>33</sup>

Pero el evolucionismo de Dennett tiene un aspecto que parece excesivo y que valdría la pena matizar: el adaptacionismo. Antes de ver esto, debemos revisar lo que nuestro autor llama la Actitud Intencional.

---

<sup>33</sup> Si entendemos al ego humano como narración, uno podría preguntarse: ¿cómo es que una narración controla algo físico (el cerebro)?, ¿no es un problema similar al del dualismo? La respuesta radica, creemos, en cómo consideramos a tal narración: ¿es acaso que una voz, un escrito, un tejido de palabras no son elementos físicos?, parece que sí (recordemos que el punto de partida de Dennett es el materialismo).

## Capítulo IV

### ***La Actitud Intencional, el adaptacionismo y la memética***

Buena parte de los argumentos de Dennett encuentra apoyo, en última instancia, en la Actitud Intencional, herramienta inocua de acuerdo con el profesor de Tufts, y que él considera una de sus aportaciones a la filosofía; ella es, además, muy cercana al adaptacionismo, como veremos; esta Actitud es ontológicamente poco comprometida con los fenómenos que trata, considerando a sus objetos de análisis, predicción y explicación, como una caja negra, pero usualmente útil en contextos cotidianos y científicos.

Frecuentemente, esta Actitud que propone el americano, es relacionada únicamente con los términos mentalistas (creer, desear, etc.) y su intento por legitimarlos, ante el hecho de que no conocemos su status ontológico; pero la Actitud Intencional tiene, además, otro propósito, muy cercano al mencionado: legitimar los “cuentos” adaptacionistas que encontramos a lo largo de los escritos del profesor de Tufts, como intentaremos mostrar.

Y en la misma línea de los “cuentos” adaptacionistas, que surgen básicamente del darwinismo universal de nuestro autor, encontramos una consecuencia fundamental de este: su aprecio por la memética, teoría poco desarrollada y que creemos deficiente para dar cuenta de la evolución cultural.

### ***Acerca de la intencionalidad***

Partamos de dos preguntas: ¿Cómo puede algo dentro de nuestra cabeza apuntar una flecha abstracta hacia una cosa dentro del mundo?, ¿cómo es posible que un determinado estado e evento cerebral represente una característica del mundo y no otra? (Dennett, 1991: 192).

La intencionalidad es, nos dice el americano, *acerquidad* (*aboutness*), algunas cosas del mundo tienen la propiedad de ser “acerca” de otras cosas, estados y sucesos, es decir, *señalan* hacia otras cosas. Recordemos que, según la tesis de irreductibilidad atribuida a Brentano, la intencionalidad es la característica propia de lo mental, pues todo evento mental exhibe intencionalidad y nada físico lo hace. El mundo se divide en dos: lo mental y lo no mental, donde lo primero no puede ser reducido a lo segundo.

Pero, análogamente –y este rasgo es aprovechado por Dennett– lo mental es catalogable como intencional, es decir, se pueden interpretar sistemas intencionales (un sistema intencional es aquel sistema cuya conducta es predecible vía la estrategia intencional [Dennett, 1989: 15]) para reducir y definir un dominio de entidades catalogables dentro de la intencionalidad, señalando así el vínculo con las ciencias cognitivas, esto es lo que nuestro autor pretende. Dicho de otro modo: podemos atribuir intencionalidad a distintos seres, que serán objeto de interpretación –con la Actitud Intencional en un primer momento y, después, con la psicología cognitiva sub-personal–, y de análisis científico. Hablaremos un poco más acerca de la Actitud Intencional y de la psicología cognitiva sub-personal un poco más adelante, en este mismo capítulo.

El tema de la intencionalidad es, como sabemos, difícil, por lo que Dennett propone extraer dos puntos acerca tal tema, que servirán para sus propósitos; el *primer punto* es que, cuando hablamos de intencionalidad, hablamos de un rasgo lógico peculiar: la opacidad referencial; se trata del principio que señala que una rosa olería igual si se llamara de otro modo.

No importa cómo nombres a algo, ese algo está ahí, con características independientes de los nombres que puedan dársele. De modo que debido a la opacidad referencial que se utiliza en la jerga intencional, podemos confiar en cierto conjunto de modismos o términos para referirnos a ese algo del que los

modismos hablan. Estos términos son “cree que”, “sabe que”, “espera que”, “entiende que”, “desea que”, etc. Este primer punto es una ventaja clave que tenemos como apoyo para la Actitud Intencional (*Intentional Stance* ó AI), como veremos, ésta trata con los términos mentalistas al interior de una “caja negra”, pues, hasta ahora, no hemos hallado los correlatos neuronales de aquéllos.

El segundo punto es que el uso de los términos intencionales *presupone la racionalidad*<sup>34</sup> *del agente, ser o sistema* al que se le atribuyen estados referidos por los términos mentalistas, por lo que la estrategia no pretende ser fundamento para la psicología. Estos dos puntos son el lugar de partida para la AI de Dennett: hay consenso –popular y científico– para hablar de la caja negra que es lo mental hasta este momento, cosa que podemos aprovechar.

Este concepto de sistema intencional lleva una carga muy pesada. Nada más y nada menos que conectar el dominio de lo intencional (creencias, deseos, nuestro universo de sentido común acerca del mundo y las acciones) y el dominio de lo no-intencional de las ciencias físicas: neuromoduladores, regiones cerebrales especializadas, impulsos nerviosos, movimientos sacádicos, etc.; es decir, nos ayuda a *hablar de la mente*.<sup>35</sup> El término, hay que señalarlo, se ha cargado con pesos aún mayores, el sostener que la intencionalidad sea una división fundamental:

[I]s a lot to expect of one concept, but nothing less than Brentano himself expected when [...] he proposed intentionality as the mark that sunders the universe in the most fundamental way: dividing the mental from the physical. (Dennett, 1981: 22).

---

<sup>34</sup> ¿Qué entiende Dennett por ‘racionalidad’?, nuestro autor nos dice: *What about the rationality one attributes to an intentional system? One starts with the ideal of perfect rationality and revises downward as circumstances dictate. That is, one starts with the assumption that people believe all the implications of their beliefs and believe no contradictory pairs of beliefs. This does not create a practical problem of clutter (infinitely many implications, for instance), for one is interested only in ensuring that the system one is predicting is rational enough to get to the particular implications that are relevant to its behavioral predicament of the moment.*(Dennett, 1989: 21).

<sup>35</sup> Recordemos que Dennett parte de un cierto tipo de materialismo no reduccionista (la psicología cognitiva sub-personal, por ejemplo, no sustituirá a la AI).

La teoría dennettiana acerca de los rasgos de la intencionalidad y sus actitudes derivadas servirá entonces para tender un puente explicativo. Esta es la estrategia que explica y predice las acciones de distintos seres y cosas. Hay, sin embargo, dos niveles más de explicación y predicción respecto de distintas cosas en el mundo: la Actitud Física y la Actitud de Diseño.

Para predecir el comportamiento de un sistema desde la primera estrategia, la física, hace falta conocer su constitución y las alteraciones que sufre en las condiciones a predecir, utilizando las leyes de la física o cualesquiera leyes de la naturaleza que conozcamos, por simples que sean. Esta estrategia es utilizada desde el químico que trabaja con materiales peligrosos hasta la cocinera que sabe cuánto tiempo dejar una olla en el fuego.

Cuando el objeto a predecir es demasiado complejo para la estrategia física –o si desconocemos las leyes de la física, o simplemente la estrategia no arroja los frutos esperados–, puede recurrirse a otra actitud, que consiste en predecir cómo ha de comportarse un objeto de acuerdo con su diseño. Por ejemplo, aunque muchas personas no tenemos idea de los principios físicos de una computadora, conocemos su uso porque sabemos para qué ha sido diseñada. Una aplicación más simple de esta estrategia es cuando predecimos cuándo sonará un despertador. Ésta es la estrategia o Actitud de Diseño.

Hay que mencionar que, para predecir el comportamiento de los objetos, los animales no-humanos o humanos podemos “brincar” de una estrategia a la otra. No hay barreras rígidas, lo que importa es cuál funciona (explica y predice) mejor.

La tercera estrategia, la que mejores resultados nos da respecto de la explicación y predicción de los humanos y de otros animales, incluso de los mecanismos evolutivos, es la AI<sup>36</sup>. Debemos primero hacer notar de dónde surge esta

---

<sup>36</sup> Los criterios para decidir cuál estrategia funciona mejor son pragmáticos: un químico en el laboratorio, por ejemplo, puede predecir y explicar la conducta de un material exótico utilizando la

estrategia o actitud. Proviene fundamentalmente de lo que Dennett llama la psicología popular<sup>37</sup>, es algo que aprendemos desde niños y que mejora con la práctica: actuar *como si* los otros tuvieran creencias, deseos y actuaran en consecuencia; esto equivale a atribuirles “optimalidad” de diseño<sup>38</sup>, sin necesidad de comprobarlo. Esto funciona en la mayoría de los casos, como vemos todos los días en nuestras relaciones con los otros. De hecho, todo el tiempo estamos usando la psicología popular para predecir y explicar conductas y pensamientos de nuestros congéneres.

### ***La Actitud Intencional***

Como ya mencionamos, esta estrategia consiste en tratar al objeto o sistema que se quiere predecir como óptimamente diseñado: como racional, con creencias, deseos, dolores, expectativas, etc. Semejante estrategia nos permite enseguida deducir qué creencias debiera tener ese agente sobre la base de su posición en el contexto (se supone, por ejemplo, lo que debiera creer para su buen desempeño en el mundo, lo que nos permite concederle objetivos); luego, siguiendo tales suposiciones se señala qué deseos tendría y por último se predice que actuará necesariamente de acuerdo a los objetivos imputados: *attribute as beliefs all the truths relevant to the system's interests (or desires) that the system's experience to date has made available* (Dennett, 1987: 18).

Con la AI Dennett busca legitimar los presupuestos que usamos cotidianamente: atribuir mentes, intenciones, diseño óptimo, etc.

---

estrategia física, pero esta estrategia no funcionaría para explicar porqué Pedro no saltó al precipicio frente a él y viceversa.

<sup>37</sup> Para Dennett la psicología popular *might [...] be viewed as a rationalistic calculus of interpretation and prediction –an idealizing, abstract, instrumentalistic interpretation method that has evolved because it works and works because we have evolved* (Dennett, 1989: 48).

<sup>38</sup> Hablaremos de diseño al referirnos a la AI y a sus predicciones y explicaciones, mientras que escribiremos “diseño” cuando hagamos referencia a los rasgos actuales de las especies.

La estrategia brinda un poder predictivo que normalmente no se puede obtener por ningún otro medio, lo cual es su punto fuerte, más que saber si los términos, correlatos físicos del cerebro o suposiciones acerca de la evolución de alguna especie, son verdaderos o no. Se trata de una muleta práctica, inocua, una poderosa antropomorfización. Se limita a la conducta interna y externa de los sujetos, pero incluye a los términos mentalistas o intencionales para clasificarlas. Con este método de caja negra los problemas de si realmente hay creencias y deseos se aplazan. No hay compromiso ontológico acerca de los términos que la estrategia supone.

Se trata entonces de atribuir *those beliefs the system ought to have*. (Dennett, 1987: 20). Esta regla implica que desde la AI se atribuyan deseos y creencias dependientes de una lista de deseos –o de supuestos– básicos como la supervivencia, evitar el dolor, la alimentación, la reproducción, el entretenimiento, la comodidad, etc., sin que tal lista sea rígida.

Por otro lado, no podemos atribuir a un sistema el deseo de autodestruirse o hacerse daño en situaciones normales; aunque sabemos que pasan cosas así, este tipo de acciones también pueden ser explicadas desde la AI, con ligeras modificaciones: cuando se atribuyen creencias falsas, la historia que se cuenta es especial para explicar cómo el sistema llegó a tales conclusiones (o cometió tal o cual error) en su relación con el mundo que lo provee de información. Se pueden atribuir deseos “anormales” bajo la misma óptica (historias y circunstancias particulares). Todo esto se encuentra también dentro de la estrategia como uno de sus matices.

La racionalidad que se atribuye a un sistema intencional, al intentar predecir su conducta, es la racionalidad ideal, perfecta; esta suposición de racionalidad se va modificando de acuerdo con las circunstancias, es decir, la estrategia no es un listado rígido de normas a seguir a modo de un manual, sino que observa

constantemente el nivel empírico para *modificar y adaptar* sus presupuestos, porque, después de todo, lo que se quiere predecir es algo complejo y cambiante. Señalaremos un poco más a este respecto cuando veamos el ejemplo de los vervets.

Aunque claro, la AI no es perfecta. Ante algunas creencias contradictorias o autoengaños de los humanos, no es tan claro qué deseos han de atribuirse en cada caso, por no mencionar lo que sucede cuando se aplica a máquinas mientras estas funcionan mal. Entonces ejercemos la Actitud de Diseño, por ejemplo cuando hay personas con raras patologías, como el delirio de Capgras o la prosopagnosia<sup>39</sup>.

Ninguna interpretación intencional funciona a la perfección, debido a que puede haber dos interpretaciones intencionales de una misma actitud o distintas predicciones en relación con un sistema; esto es, la presencia de un modelo no excluye la existencia de otro modelo aplicado a la misma situación, aunque esta posibilidad es insignificante en relación a los beneficios que la AI nos aporta; en efecto, el riesgo que se corre de cometer errores usando esta presunción de lo óptimo bien vale la pena a cambio de su velocidad y eficiencia en situaciones cotidianas.

Además, vale la pena el riesgo porque se trata de sistemas intencionales complejos, y las predicciones y explicaciones pueden ir modificándose: se agregan datos de la relación del sistema con el entorno hasta llegar a una única interpretación sobre la base de que el sistema tiene estados informados acerca del mundo: las preguntas (y predicciones por lo tanto) se van haciendo específicas; por ejemplo: “¿qué cree usted acerca de la caldera?”; “que está apagada” “¿De

---

<sup>39</sup> Quienes padecen de la patología llamada delirio de Capgras creen que un ser querido ha sido reemplazado por un doble. En la segunda patología, la prosopagnosia, el enfermo puede distinguir a una mujer de un hombre, a un africano de un asiático, etc., pero no puede reconocer a sus familiares o amigos sino hasta que escucha su voz.

cuál caldera me habla?"; "de la que está en el sótano", etc.; esto porque el organismo refleja continuamente el entorno.

La pregunta difícil, en todo caso, es: ¿por qué funciona esta estrategia (como podemos comprobarlo en nuestra actividad cotidiana)?

Esto es, si la AI es actuar como si el hombre, máquina, mecanismo, sistema o cosa es actuar *como si* tuviera *x* creencia, o actuara de acuerdo a ciertos fines, lo natural sería que el paso siguiente fuera descubrir si *realmente* esas creencias y deseos atribuidos están dentro del sistema intencional; es decir, no se trataría de descubrir si la estrategia predictiva funciona, sino de descubrir *–saber–* si realmente el objeto tiene esas creencias, lo cual implicaría conocer antes el status ontológico de las creencias.

La primera respuesta de nuestro autor acerca del porqué del funcionamiento de la AI recurre a "cuentos" evolucionistas, como en el caso de la evolución de la conciencia y del ego: la evolución ha diseñado a los seres humanos para ser racionales, para creer lo que deben creer y desear lo que deben desear y así sobrevivir. Como productos de un largo y exigente proceso evolutivo podemos tener confianza en que es relativamente seguro utilizar la AI.

La respuesta evolucionista es una respuesta relativamente simple: suponemos que el "diseño" que observamos en la naturaleza es el óptimo, esto es, creemos que los animales (humanos y no-humanos) actuarán de acuerdo con metas que han sido puestas por la evolución, por ejemplo, que tratarán de evadir el dolor, que buscarán comodidad y alimento, etc.

Pero insistimos, ¿cómo funciona el mecanismo con el que la evolución nos ha dotado? Esta es una pregunta sin respuesta hasta ahora. Sabemos cómo funciona la estrategia, pero no por qué funciona, sólo podemos suponerlo.

A pesar de los errores que lleguen a cometerse, a este respecto se puede sostener una actitud optimista hasta el final: desde explicar el surgimiento del “diseño” de los ojos hasta tramar estrategias pasando por las creencias y los deseos. La AI no imputa ningún poder oculto al sistema cuya conducta se quiere predecir, es metafísicamente inocuo, ontológicamente poco comprometido, pero práctico. Sólo imputa lo que el sentido común atribuye a la evolución y al aprendizaje.

Hasta aquí podemos notar que la psicología popular se parece mucho a la AI, parece que hablamos de lo mismo. ¿Cuáles son las diferencias? la diferencia clave es, como mencionamos, que la AI hace *explícito* su conocimiento, busca legitimarlo<sup>40</sup>, reconoce como caja negra al sistema observado y observable sólo en cuanto a su conducta, utiliza los términos mentalistas como técnicos, como instrumentos para explicar y predecir, sin comprometerse con su status ontológico. Adoptar la AI *no es cambiar de método* (el de la psicología popular) sino hacer explícito lo que logramos tácitamente.<sup>41</sup>

Podemos atribuir optimalidad a la naturaleza y creencias a las personas, animales no-humanos y cosas del mismo modo que hablamos de centros de gravedad y del ecuador, es decir, son ficciones útiles (Dennett, 1989: 72).

Ahora bien, Dennett nos dice que podemos confiar en la AI, pero hallaremos en el futuro su complemento científico: la psicología cognitiva sub-personal (sub-personal cognitive psychology), ciencia futura que habrá de encargarse de

---

<sup>40</sup> Este punto puede verse con el método de la tercera persona propuesto por Dennett para tratar con los términos mentalistas, que él llama heterofenomenología. Hemos mencionado en la introducción (Cf. p. v) que esta parte de la postura filosófica de Dennett no está dentro de los límites de este trabajo, tal discusión requeriría un tratamiento detallado, pues la pretensión del profesor de Tufts parece ser, en última instancia, que la heterofenomenología sustituya a la fenomenología para el estudio de la mente.

<sup>41</sup> Creemos que Dennett, al mencionar que no es necesario un “cambio de método”, se refiere al proyecto reduccionista de Paul y Patricia Churchland, que consiste en probar que la psicología popular es una teoría y, además, mostrar porqué es una mala teoría y ha de ser desechada. Con este punto se puede notar, creemos, cómo Dennett se aleja del materialismo eliminativo.

demostrar las presiones sobre el “diseño” del cerebro y cómo se implementan los sistemas que hacen posible la AI; es decir, debe mostrar cómo y por qué tenemos éxito en comprendernos como sistemas intencionales.

La AI será parcialmente desmentida una vez eliminados algunos de los supuestos –reconocidos desde el principio como vulnerables– con los que nos conducimos en la vida cotidiana; el concepto de creencia estará en manos de la psicología cognitiva sub-personal para su explicación microfísica, concreta, microteórica, pero seguirá hallando en la AI su respuesta económica, abstracta, holística e instrumental: éste es el “divorcio” propuesto por Dennett respecto de las creencias (Dennett, 1987: 57).

El status de las creencias de la AI, pues, se mantendrá como hasta ahora, mientras que la psicología cognitiva sub-personal hará lo propio; es decir, explicar científicamente qué son los términos mentalistas. Pero aun entonces, el principio de economía que representa la AI seguirá siendo la opción más viable para el uso cotidiano.

Se han de proponer y probar modelos acerca de cómo el cerebro hace lo que hace, cómo se implementan pautas para el aprendizaje de conceptos, expectativas, solución de problemas, conductas dirigidas hacia fines determinados, determinación de fines, extrapolaciones, manejo de herramientas y combinaciones conceptuales, etc., que no sólo son manejo de sensaciones y respuestas de tropismo, sino que producen un simulacro de sensibilidad genuina del contenido, es decir, semántica, sin abandonar la AI.

Así como la física y la química se entretrejieron, nos dice Dennett, el progreso de la psicología cognitiva sub-personal borrarán los límites entre ella y la AI, toda vez que se encuentren los correlatos neuronales de los términos mentalistas<sup>42</sup>; más aun, la

---

<sup>42</sup> Con afirmaciones como ésta podemos ver el realismo de Dennett, que le resulta problemático, particularmente respecto de las explicaciones evolucionistas que propone para la mente, sujetas a su

psicología debe encaminarse, siguiendo esa vía, a señalar los fundamentos fisiológicos de toda conducta (psicológica) sin eliminar sus caracterizaciones intencionales, así como a destacar la manera cómo el cerebro parece extraer semántica de la sintaxis. El mejor candidato para esta tarea de psicología es, de acuerdo con Dennett, la neurofisiología.

Pero, ¿la AI podrá reducirse a la psicología cognitiva sub-personal, eliminando así las caracterizaciones intencionales?; la respuesta dennettiana es que no, por lo expuesto: podemos utilizar la AI para seguir comprendiéndonos, pero utilizándola con cautela, sin tomar por reales los términos que parecen hablar del interior de la caja negra.

Volviendo a la pregunta inicial de este apartado, ¿cómo es que algo apunta hacia el exterior?; ¿cuál es esa “flecha” en el cerebro que señala hacia el mundo? Bueno, lo sabremos al identificar los elementos fisiológicos con la semántica del sistema.

If we can [...] legitimize "mentalistic" talk, we will have no need of a reduction, and that is the point of the concept of an intentional system [...] the final reductive task would be to show not how the terms of intentional system theory are eliminable in favor of physiological terms via sub-personal cognitive psychology, but almost the reverse: to show how a system described in physiological terms could warrant an interpretation as a realized intentional system (Dennett, 1987: 67-68).

Ya hemos abundado en que la AI y su intento por legitimar los términos mentalistas, sin comprometerse ontológicamente con su existencia, es una *herramienta* para predecir y explicar animales humanos y no-humanos, sistemas, objetos, etc., y no una teoría, sino una herramienta<sup>43</sup> (Dennett habla de la Teoría de los Sistemas Intencionales, pero tal teoría es sólo es suponer, con bases prácticas, que la AI funciona [Dennett, 2009b: 7-8]).

---

adaptacionismo. Veremos esto cuando mencionemos las críticas de Gould y Lewontin al programa adaptacionista, en el último capítulo.

<sup>43</sup> Recordemos que la AI es, hasta este momento, una herramienta para predecir y explicar la conducta de otros seres, pero desconocemos cómo funciona al interior de la “caja negra”, algo que nos dirá la psicología cognitiva sub-personal.

### ***Defensa del adaptacionismo***

Ahora bien, la AI, como hemos visto, también se apoya en los “cuentos” evolucionistas mencionados acerca de la conciencia y del ego, pues suponen el “diseño” óptimo en la naturaleza. Podemos decir que la AI tiene una doble función: legitimar a los términos mentalistas y, por otro lado, legitimar a los “cuentos” evolucionistas.

La confianza en la optimalidad que muestra la AI supone, como podemos ver, el adaptacionismo o Paradigma Panglossiano del que Stephen Jay Gould y Richard Lewontin hablan, nombrado así por la caricatura que Voltaire hace de Leibniz en la novela *Candide*, donde el Dr. Pangloss sostiene que este es el mejor de los mundos posibles. El punto es que nada, en principio, podría probar lo contrario.

Para los autores de Harvard, el adaptacionismo [is a] *programme based on the faith in the power of natural selection as an optimizing agent* (Gould & Lewontin, 1979: 1<sup>44</sup>; Dennett, 1989: 260).

La crítica hecha por los estudiosos de Harvard a los “programas adaptacionistas” es justa, y tiene puntos importantes que nos recuerdan la cautela necesaria cuando hablamos de “cuentos” evolucionistas y erigimos mecanismos que parecen demasiado simples, además de irrefutables.

Aunque la interpretación del texto de Gould y Lewontin está sujeta a muchas discusiones, podemos aceptar como puntos de su argumento anti-adaptacionista a los siguientes:

1. Las afirmaciones del adaptacionismo son difícilmente falseables (algunas son imposibles de falsear, añadiríamos nosotros).<sup>45</sup>

---

<sup>44</sup> Recurso electrónico en lenguaje HTML, paginación de Word fuente 12.

2. Parece demasiado fácil crear “cuentos” evolucionistas. Siempre se puede contar con una historia o cuento y alegar ignorancia temporal en las fallas o lagunas del cuento y tener confianza en el futuro científico.<sup>46</sup>
3. Ver en la selección natural el mecanismo central de la evolución desvía de otros detalles importantes como la deriva génica y los constreñimientos de arquitectura: *optimality arguments dispense with the tedious necessity of knowing anything concrete about the genetic basis of evolution* (Lewontin, 1978: 14)<sup>47</sup>.

Por lo anterior, el adaptacionismo falla en distinguir la utilidad actual de los rasgos de una especie con sus razones de origen; esto se ve, por ejemplo, con la exaptación<sup>48</sup>: rasgos actuales útiles que surgen de estructuras utilizadas previamente con otros fines, por ejemplo, el pulgar del panda menor, previamente utilizado para desplazarse en los árboles, usado después para manejar su alimento habitual, que es el bambú.

Otro ejemplo que nos puede ayudar a mostrar este punto es el de los zorros domesticados por Dmitri Belyaev a partir de 1959: la selección reproductiva favoreció a los zorros menos agresivos, pero después de diez generaciones, además de un temperamento relativamente dócil, los zorros mostraban conductas muy distintas a los zorros salvajes, más parecidas a las de los perros domésticos, además de pelaje de coloración distinta y hocico achatado.

---

<sup>45</sup> *We would not object so strenuously to the adaptationist programme if its invocation, in any particular case, could lead in principle to its rejection for want of evidence. We might still view it as restrictive and object to its status as an argument of first choice. But if it could be dismissed after failing some explicit test, then alternatives would get their chance* (Gould & Lewontin, 1979: 5).

<sup>46</sup> *Since the range of adaptive stories is as wide as our minds are fertile, new stones can always be postulated. And if a story is not immediately available, one can always plead temporary ignorance and trust that it will be forthcoming* (Gould & Lewontin, 1979: 5).

<sup>47</sup> Texto en lenguaje HTML, paginación de Word fuente 12.

<sup>48</sup> El concepto es acuñado por Gould y Elisabeth S. Vrba en el artículo de 1982 llamado *Exaptation: a missing term in the science of form*.

Ésta es, precisamente, la idea que ilustran los *spandrels* en la arquitectura, en donde Gould y Lewontin reflejan su postura: subproductos, no planeados originalmente, pero definidos por la necesidad que la construcción pueda contener a la vez cúpulas y arcos.

Como Dennett mismo señala, la crítica de los estudiosos de Harvard recuerda a B. F. Skinner, en su crítica al mentalismo, donde las vías de explicación mediante atribución de creencias y de deseos son muy fáciles, y distraen al psicólogo de los detalles importantes: la historia de reforzamiento (Dennett, 1987: 262). Esta crítica se parece también, recordemos, a la de Karl Popper a la psicología individual de Adler y al psicoanálisis de Freud, acusándolas de ser afirmaciones imposibles de someter a falsación.

La AI, como dijimos, sirve no sólo para los términos mentalistas, sino para distintas teorías que pueden ser una especie de caja negra también, precisamente como el desarrollo de los rasgos de alguna especie; más aun, los problemas de interpretación en psicología y los problemas de interpretación en biología *are the same problems, engendering the same prospects –and false hopes– of solution, the same confusions, the same criticisms and arguments. That is a voluminous claim, and there is a tendency to underestimate its ramifications* (Dennett, 1987: 277).

La evolución de las especies es otra caja negra, en el sentido de que es imposible saber la historia a detalle de los rasgos actuales que observamos en la naturaleza. El mismo problema que tenía el conductismo, respecto de la imposibilidad práctica de hallar la historia de reforzamiento que impulsaba una conducta, se presenta en las explicaciones evolucionistas. Queremos saber por qué, pero es imposible conocer *la historia completa*, en detalle: ¿sugiere el programa anti-adaptacionista esta vía sincrónica? Parece que no:

The contrast between the two shorts of answers, the scrupulously non-adaptationist historic-architectural answer Gould and Lewontin *seem* to be championing and the frankly Panglossian adaptationist answer one can also try to give, is vividly captured in one final analogy from the Skinnerian war against mentalism. (Dennett, 1987: 267).

Podemos decir es que el adaptacionismo, según los autores de Harvard, debe ser reemplazado por una especie de pluralismo (Gould & Lewontin, 1979: 9) en el pensamiento biológico, donde se consideren otros aspectos importantes que *pueden* haber influido en los rasgos actuales de una especie. ¿Cumple el adaptacionismo de Dennett con esto?, parece que no, aunque hay que hacer matices importantes respecto de la postura del profesor de Tufts.

La táctica de la AI no es reemplazar a las investigaciones empíricas por investigaciones apriorísticas, sino de usarla para sugerir qué preguntas empíricas han de hacerse a la naturaleza. Podemos verificar las hipótesis en competencia explotando la presunción de racionalidad de la AI. Esto es, los “cuentos” evolucionistas han de ser puestos a prueba, evitando la excesiva confianza en ellos, como veremos en casos prácticos con animales no humanos.

Recordemos que Dennett nos habla de que un error fundamental del sentido común, al pensar en la conciencia, es que no distingue entre la forma reflexiva del pensamiento (donde  $x \neq y$ ) y la no reflexiva (donde  $x = y$ ), identificando así un estado consciente con el tener meta-creencias que implican el TC o lugar central para el observador. Pues bien, el papel de las meta-creencias (creencias de primer, segundo, tercer orden...) nos ayuda nuevamente como pauta para la investigación empírica.

En el caso de la suposición de optimalidad en el “diseño” de una especie (por ejemplo, el caso de los vervets que veremos más abajo), y teniendo en mente que la AI se guía por la reflexividad, podemos preguntarnos ¿cuál nivel de pensamiento se les puede atribuir a los animales no-humanos, dado que no podemos hablar con ellos?; ¿de primer orden, de segundo orden, etc.?, y esto con el fin de ir eliminando candidatos a respuesta (el punto de partida, nuevamente, es

el sentido común: un etólogo difícilmente partirá de asumir creencias de cuarto orden en los micos, por ejemplo):

We can test the competing hypotheses by exploiting the rationality assumption of the intentional stance. We can start at either end of the spectrum; either casting about for the depressing sorts of evidence that will *demote* a creature from a high-order interpretation, or hunting for the delighting sorts of evidence that *promote* creatures to higher-order interpretations (Dennett, 1987: 247).

Podemos, por ejemplo, atribuir cero orden a los monos vervet, y considerar que el gruñido que emiten ante el peligro de depredadores (*Moving Into the Open* ó MIO) es simplemente una señal de ansiedad, y que “realmente” no saben o “no tienen conciencia” de que el sonido que emiten funciona como alerta para los otros monos, sin embargo

We are delighted to learn, for instance, that lone male vervet monkeys, traveling between bands (and hence out of the hearing, so far as they know, of other vervets) will, on seeing a leopard, *silently* seek refuge in the trees. So much for the killjoy hypothesis about leopard-anxiety yelps. (Dennett, 1987: 247).

Vemos, pues, cómo la AI, mediante la atribución de creencias y la optimalidad, puede ayudar a la investigación empírica, pero no sólo respecto de la conducta observada, sino también para otros rasgos funcionales; veamos el ejemplo de los zopilotes aura, que son atraídos por la carroña: parece que algo tan repulsivo para los humanos, es atractivo para ellos; tal vez podemos extrapolar nuestro caso: los zopilotes aura muestran una conducta impaciente al abalanzarse sobre la carroña, podríamos decir que *les gusta* en un sentido parecido en el que a los humanos nos gusta algún platillo. Pero podría decirse que:

[y]ou are being awfully sloppy in these everyday attributions. Vulture revulsion is possible [...] you would not make their observed behavior criterial of pleasure, would you? The suggestion that it makes no sense for vultures to be disgusted by their designated diet is nothing but Panglossian optimism (Dennett, 1998: 341).

Tal vez los ancestros de los zopilotes aura se hallaron en un callejón sin salida: odiaban el sabor de la carroña, pero al ser la única comida disponible, tuvieron que superar su disgusto y consumirla y, tal vez, desde entonces, desarrollaron una

especie de comportamiento estoico, de modo que, lo que interpretamos como gusto por la carroña, es en realidad desesperación.

My rush to judgment was perhaps a bit rash, so let's explore further to see whether any supporting evidence can be found for your alternative hypothesis. [...] It would be fascinating to discover something along the lines of an incompletely suppressed gag-reflex as part of the normal vulture feeding behavior, or perhaps some traces of approach-avoidance opponent systems tugging away at each other in their brains, a sort of activity not to be found, we might imagine, in the brains of birds with more savory diets [...] [oversimplified] interpretations would be overthrown by more sophisticated use of behavioral considerations [...] This might be –indeed ought to be– just the beginning of a long and intricate examination of the possible functional interpretations of events in vultures' nervous systems, (Dennett, 1998: 342).

Así, la atribución de optimalidad, junto con la AI, nos puede ayudar en la investigación empírica, pero no reemplazándola. Pero volvamos a lo que mencionamos antes, ¿muestra Dennett esa cautela cuanto urde su cuento acerca de la conciencia, por ejemplo?, parece que no, pues no considera ningún aspecto de los señalados en la crítica que hacen Gould y Lewontin al programa adaptacionista; entendemos que lo importante de las metáforas del americano es mostrar su punto –ayudar a pensar algunos fenómenos, como el ego, en su búsqueda por naturalizarlos–, pero podemos decir que en ninguna de las metáforas que ha usado Dennett, hasta ahora, se muestra un cuidado como el que él mismo recomienda a los etólogos y biólogos.

En los “cuentos” evolucionistas, nuestro autor parece dar intentos de explicación demasiado sencillos –monocausales– en relación a lo que parece un asunto más complejo, como el surgimiento de la conciencia refleja.

Ahora bien, ¿así como la psicología cognitiva sub-personal complementará a la AI, explicándola (mostrando los correlatos neuronales de los términos mentalistas), podemos decir que la biología complementará a los “cuentos” evolucionistas?; tal vez, pero hay que tener paciencia. Pero, esta es, precisamente, una de las críticas que se le pueden hacer al adaptacionismo, recordemos, es: que frecuentemente se hacen “cuentos” irrefutables y a la vez se tiene esperanza en el poder

explicativo de la ciencia futura (como en el caso del “cuento” que Dennett urde respecto del surgimiento de la conciencia).

De modo que en los “cuentos” mostrados, Dennett no utiliza aquello que le parece una buena idea: el pluralismo en las explicaciones evolucionistas (Dennett, 1987: 261), aun cuando son pasos sumamente importantes en su discurso. Urdir “cuentos” acerca de especies que podemos observar, y elaborar hipótesis que pueden ser puestas a prueba –como en el caso de los vervets o los zopilotes aura–, es algo que parece un tanto disímil en relación con el surgimiento de la conciencia refleja. El adaptacionismo Dennettiano, entonces, debe ser visto como un buen esfuerzo –uno pionero–, siempre y cuando, –él mismo lo advierte respecto de la Actitud Intencional–, se sigan sus hipótesis con cautela.

Es en este punto donde la postura dennettiana, que aparece como algo razonable, comienza a flaquear: Dennett lleva su darwinismo universal un tanto lejos, pues si bien podemos predecir y explicar distintos fenómenos, la misma actitud, extrapolada, conduce a nuestro autor a suponer una serie de metáforas que se toman en sentido literal, peligro que el mismo americano nos había aconsejado prevenir (Dennett, 1991: 455). La teoría de los memes, al contrario de las metáforas que nuestro autor propone para pensar la conciencia, se muestra incapaz de encontrar apoyo en la ciencia actual<sup>49</sup>. Esta actitud es vista con la

---

<sup>49</sup> Creemos que buena parte de la postura de Dennett, aunque rara vez cita directamente a neurocientíficos, proviene, precisamente, de las neurociencias: de acuerdo con Damasio, por ejemplo, el “error de Descartes” consiste en ignorar que la Naturaleza *appears to have built the apparatus of rationality not just on top of the apparatus of biological regulation, but also from it and with it* (1994: 128), de modo que ha de buscarse la vía evolucionista para la explicación de los fenómenos mentales, evitando ver a la conciencia como un fenómeno que surge desde una región cerebral como fuente primaria (como parece que ocurre con algunas interpretaciones de los experimentos de Libet). Otro aspecto, también señalado por Damasio, es que valdría la pena evitar hablar de *qualia*, para considerar al organismo entero en la génesis de la subjetividad –algo que Dennett también recomienda, aunque no lo apuntamos en este trabajo por motivos de espacio– (Damasio, 1994: 224-225). Que un rasgo del cerebro es su multifuncionalidad, es apoyado por los estudios de Walter J. Freeman (2000); Gazzaniga (1998: 128) y Damasio (2000: 174), sostienen que el ego humano es una especie de narración. El punto es que las metáforas de Dennett se apoyan en las ciencias cognitivas, pero de una manera crítica. A pesar de que valdría la pena abundar en este tema, pues las similitudes de la propuesta filosófica de Dennett y los estudios de algunos científicos

teoría memética, de la cual Dennett es un férreo partidario; de acuerdo con ella, los memes son la más reciente expresión de la evolución.

### ***La teoría memética***

En el libro llamado *The Selfish Gene* de 1976, Richard Dawkins introdujo el término *meme*, para referirse a lo que él considera las unidades de transmisión cultural que han surgido recientemente, obedeciendo a *todas* las leyes de la evolución y de un modo análogo a los genes. En el libro mencionado, la discusión se centraba en torno al objeto de la evolución: ¿eran los fenotipos o los genes las unidades seleccionadas por la naturaleza?

Para Dawkins, son los genes. Los seres humanos, entonces, son “simplemente” vehículos de genes; éstos sólo tienen una “misión”, a saber, replicarse de acuerdo a su propio beneficio; desde esta perspectiva, podemos decir lo mismo acerca de los memes:

What, after all, is so special about genes? The answer is that they are replicators [...] I think that a new kind of replicator has recently emerged on this very planet. It is staring us in the face.

The new soup is the soup of human culture. We need a name for the new replicator, a noun that conveys the idea of a unit of cultural transmission, or a unit of imitation. I want a monosyllable that sounds a bit like 'gene'. I hope my classicist friends will forgive me if I abbreviate *mimeme* to meme. If it is any consolation, it could alternatively be thought of as being related to 'memory', or to the French word meme.

Examples of memes are tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperms or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain via a process which, in the broad sense, can be called imitation. If a scientist hears, or reads about, a good idea, he passes it on to his colleagues and students. He mentions it in his articles and his lectures. If the idea catches on, it can be said to propagate itself, spreading from brain to brain. (Dawkins, 1976: 213).

---

reconocidos pueden ser objeto de importantes matices, no está dentro de los límites de este trabajo, tal empresa requeriría un trabajo independiente.

Esta postura es sostenida por Dawkins en sus siguientes libros, por ejemplo, en *The Extended Phenotype* (1982). Para este norteamericano, el punto de vista de los memes para el estudio de la cultura (memes-eye-viewpoint) es sumamente fructífero. Tiempo después parece que Dawkins modificó o suavizó sus afirmaciones, haciendo divisiones entre tipos de memes –de “alta fidelidad” y de “baja fidelidad”– por ejemplo, tratando así de responder a la objeción de que la difusión de las unidades de transmisión cultural es distinta a la de los genes, pues la imitación –el modo en que se transmiten los memes– no garantiza una “copia fiel”.

Así, las unidades de transmisión cultural de “alta fidelidad” son los que podemos comparar con los genes. Por ejemplo: los niños imitan las letras que observan a los adultos escribir, pero sus trazos distan mucho de ser copias exactas (son memes de “baja fidelidad”); pero si ya adquirieron el meme de la escritura, los trazos de los infantes son más fieles, y, por lo tanto, el propio meme que se está replicando (Dawkins, 2003: 121). Es decir, el rango de mutación y de variabilidad de los memes es más alto que el de los genes. El problema que surge es que tal vez el rango de mutación sea tan alto que difícilmente podemos hablar de que el meme emitido (por la escritura de los adultos, en el ejemplo anterior), y el que recibe el huésped humano (los niños) sean el mismo.

Si la replicación es tan rápida, es probable que fuera demasiado caótica para ser analizada en términos de un esquema científico. Dennett admite esto, sin encontrar una respuesta satisfactoria:

[W]e must be careful about a problem we have already identified in several different guises: the problem of how to tell plagiarism (or respectful borrowing) from convergent evolution [...]. We do not want to consider two *identical* cultural items as instances of the same *meme* unless they are related by descent. (The genes for octopus eyes are not the same genes as those for dolphin eyes, however similar the eyes may appear.) This is apt to create a host of illusions, or just undecidability, for cultural evolutionists whenever they attempt to trace the memes for Good Tricks. The more abstract the level at which we identify the memes, the harder it is to tell convergent evolution from descent.

The more purely semantic our principles of identification are [...] the harder it is to trace descent with confidence. This is the same epistemological problem, in the science of culture, that taxonomists confront when they try to sort out homology from analogy, ancestral from derived characters, in cladistic analysis

Dawkins, además, habla posteriormente de selección de memes altruistas (Dawkins, 2003: 133) cuando, en *The Selfish Gene*, se hace énfasis en el egoísmo de los replicantes (genes y memes). Aun así, la teoría memética sumó allegados rápidamente, entre los cuales se encuentra el propio Daniel Dennett.

¿Es la teoría una extrapolación demasiado simplista de, primero, las líneas que siguen –como la imitación– y, segundo, de la teoría de la evolución por selección natural que el mismo Dawkins sostiene? Esto, porque algunos partidarios de la teoría (como Dennett y Dawkins) sugieren que las unidades de transmisión cultural adelantan en un sentido evolutivo a los genes, pues el nuevo replicante *is still in its infancy, still drifting clumsily about in its primeval soup, but already it is achieving evolutionary change at a rate that leaves the old gene panting far behind* (Dawkins, 1976: 213). A diferencia de los genes, prácticamente cualquier cosa puede ser un meme: desde un trozo de papel hasta la teoría de la relatividad.

Aunque se sostiene que la memética obedece a todas las leyes de la evolución, a primera vista el proceso de evolución memética parece la idea lamarckista de que las características adquiridas durante la vida del fenotipo son heredadas por sus descendientes. ¿Es la evolución memética un proceso lamarckista, darwiniano, o una combinación de ambos procesos? Dennett admite, por ejemplo, que en su versión de la memética es innegable que se trata de un proceso lamarckista similar al efecto Baldwin; aunque no le parece una falla (Dennett, 1995: 355).

La ambición subyacente de la teoría memética es la de hallar leyes fundamentales de la biología, aplicables a todo el universo, a la manera de la física:

The laws of physics are supposed to be true all over the accessible universe. Are there any principles of biology that are likely to have similar universal validity? When astronauts voyage to distant planets and look for life, they can expect to find creatures

too strange and unearthly for us to imagine. But is there anything that must be true of all life, wherever it is found, and whatever the basis of its chemistry? If forms of life exist whose chemistry is based on silicon rather than carbon, or ammonia rather than water, if creatures are discovered that boil to death at -100 degrees centigrade, if a form of life is found that is not based on chemistry at all but on electronic reverberating circuits, will there still be any general principle that is true of all life? Obviously I do not know but, if I had to bet, I would put my money on one fundamental principle. This is the law that all life evolves by the differential survival of replicating entities. The gene, the DNA molecule, happens to be the replicating entity that prevails on our own planet. There may be others. If there are, provided certain other conditions are met, they will almost inevitably tend to become the basis for an evolutionary process. (Dawkins, 1976: 192).

Al interior de la teoría memética, además, podemos hacer una división entre partidarios: internistas y externistas (Dennett, 1995: 353). Los primeros, entre los que podríamos incluir a Robert Aunger (2002) y Juan Delius (1991), sostienen que un meme es localizable (al menos en principio) en el cerebro humano: es decir, tales unidades de algún modo viajan entre un cerebro y otro y se alojan en una región específica del cerebro:

Memes are capable of instructing, not protein synthesis as genes do, but behavior. However, genes can do that indirectly through protein synthesis. On the other hand meme replication, by involving neurostructural modifications, is invariably associated with the induction of protein synthesis. (Delius, 1991: 84)

Robert Aunger, en *The Electric Meme* (2002) y, como el propio título lo indica, sostiene que los memes han de ser localizados en la comunicación entre neuronas si se quiere que pierdan su status de metáfora

Something new will have to be done if memetics is going to advance and become a viable alternative to standard theories of cultural change [...] no one knows what a meme is. Certainly the existence of one has yet to be demonstrated (Aunger, 2002: 21).

Pero el primer problema que se presenta es que los impulsos eléctricos entre neuronas son demasiado veloces como para que haya tiempo de aislar e identificar allí a cualquier meme (recordemos la medición de tiempo a escala microscópica del cerebro, que traía problemas al modelo del TC). Aunger admite este problema, y sin darle solución, sólo señala la vía posible para localizar a los memes:

This doesn't mean that replication of neuronal states occurs, but it does imply that there is *room* for another replicator to be responsible for brain activity [...] However, this alternative replicator must work very fast indeed (Aunger, 2002: 192).

Aunque actualmente es posible notar que cuando un humano está llevando a cabo una tarea de apreciación estética usa una o más regiones específicas del cerebro, e incluso las áreas de producción de lenguaje son localizables, etc., no pasa lo mismo con los memes (por lo menos no hasta ahora).

¿Se podrá alguna vez localizar en cierta región cerebral la creencia, por ejemplo, en el materialismo filosófico?; ésta es otra parte ambiciosa, punto de partida de la teoría para algunos (y hasta ahora también sus límites). Refiriéndose a las mutaciones en los memes, Aunger nos dice: *Can neuronal nodes take on many states and still remain infectious? The answer to this question is much less clear [...] because we still don't know enough about the brain to say for sure* (Aunger, 2002: 215).

De modo que, aunque es posible *en principio*, hasta ahora no se ha podido localizar a los memes eléctricos, y tampoco hay buenas pistas que nos guíen hacia ellos (escribir todas las conversaciones de menos de 100 palabras realizadas hoy en una cafetería es también posible en principio, pero parece difícil de realizar en la práctica).

Los externalistas, por su parte (entre ellos Dennett), sostienen que no es tan importante localizar en el cerebro a los memes, sino que éstos importan por su función: como en el caso de los genes en el párrafo que citamos al principio (de Dawkins), lo crucial es que son replicantes, unidades de transmisión que, según el profesor de Tufts, tienen sus tres pilares en: 1.- la variabilidad memética o aparición al azar de mutaciones, 2.- la herencia y 3.- la selección natural.

Esto busca salvar la problemática de hallar la base material de las unidades de transmisión cultural e intenta explicar algunas de las objeciones “estándar” a la teoría: si bien estas unidades de transmisión presentan algunos cambios o

desajustes, éstos son explicados por la novedad del replicante, y porque cualquier cosa que se aplique a la primera se aplica a la segunda (hay presión selectiva, memósfera del mismo modo en el que hay biosfera, lucha por la supervivencia entre los memes, etc.) *pero no al revés* (Dennett, 1991:426).

Hasta ahora, a los partidarios de la memética les parece que la teoría es una buena idea y que, de hecho, es posible una ciencia de tales unidades de transmisión cultural, aunque no existan avances que señalen por dónde ha de empezar esa ciencia.

Respecto de la transmisión de los memes. ¿Qué clase de mecanismo utilizan los memes para viajar y reproducirse? La primera respuesta, como vimos, es la imitación, sostenida por Dawkins (1979: 213) y, además, por Dennett : *A wagon with spoked wheels carries not only grain or freight from place to place; it carries the brilliant idea of a wagon with spoked wheels from mind to mind* (Dennett, 1991: 204).

Sin embargo, el principal problema de la idea de que la imitación está a la base del proceso, es que aquélla no puede ser la forma básica de la acción humana, porque presupone algo para ser imitado; es decir, necesita un tipo de acción más elemental. Si el quehacer humano fuera sólo imitación, esto limitaría su conexión con el mundo circundante, lo cual sería grave, pues el ambiente es precisamente el que elige quién sobrevive (de hecho ésta es la idea de Dawkins en *The Extended Phenotype* [1982], por ello, el vehículo de genes o memes es indisoluble del ambiente). El problema con la imitación es expuesto por Peter Richerson y Robert Boyd de la siguiente manera:

Culture is adaptive because the behavior of other individuals is a rich source of information about which behaviors are adaptive and which are not. We all know that plagiarism is often easier than the hard work of doing something by ourselves; imitating the behavior of others can be adaptive for the same reason. The trick is that once culture becomes important, the nature of the behavior that is available to imitate is itself strongly affected by the psychology that shapes how we learn from others. To take an extreme example, if everyone relied completely on imitation, behavior would

become decoupled from the environment. With any environmental change, imitation would be no longer adaptive (Richerson and Boyd, 2005: 12).

Aun si concedemos que, considerado en su sentido más amplio –recordemos cuando Dawkins habla de “alta fidelidad” en la transmisión memética–, el mecanismo básico de transmisión de memes puede llamarse de imitación, ¿cuáles son los detalles de ese proceso, o los elementos restantes o mínimos?, ¿cuáles son las alternativas de replicación para los memes? Hasta ahora, pues, parece que algunos de los asuntos pendientes para la hipotética ciencia de los memes son:

1. La teoría memética se mueve todavía en el terreno de las analogías y las metáforas, al menos hasta que consiga un desarrollo cuantitativo y matematizado, y sea capaz de establecer con mayor precisión qué debemos entender propiamente cuando hablemos de “memes”, y de cómo localizarlos. Por ello, la analogía gene/meme parece débil, debido a que con el primero se tiene una base más sólida e identificable para su estudio, algo con lo que la teoría memética no cuenta.
2. No todas las ideas son replicantes, por lo tanto no todas las ideas son memes. Por ejemplo, un pastel que alguien prueba<sup>50</sup>, dispara en él la intención de hacer algo similar. El meme (la receta del pastel) no es copiado directamente, sino que remite a un conocimiento previo de X (X intentará hacer un pastel similar, aunque difícilmente el mismo). Aquí la pregunta es: ¿y de dónde vino el meme que X tiene para los pasteles? (esto es el problema de la imitación, mencionado antes).
3. No es posible rastrear el linaje de un meme, ellos no forman linajes como los genes; es decir, no es claro de dónde vino un meme en particular (aunque la noción de meme particular también puede ser puesta en duda,

---

<sup>50</sup>El ejemplo es burdo, pero Dawkins menciona al origami para hablar de cómo podemos encontrar algunos memes de “alta fidelidad” (2003: 122-123), lo cual nos hace sentir confiados con nuestro ejemplo.

pues carece de base material), debido en gran parte al “alto grado de mutación” de los memes. Esto podría responderse de la siguiente manera: precisamente, como producto más reciente de la evolución, los memes son replicantes más eficaces en este sentido que los genes, por lo que habría variación mayor. Pero si hay diferencias sustanciales entre ambos replicantes, como la antes mencionada, y si, además, el proceso de replicación de los memes aparentemente es lamarckiano, ¿hasta dónde llega la comparación válida entre ambos replicantes? ¿Es necesaria para el surgimiento y/o justificación de la teoría memética?

4. La cultura no puede “atomizarse” en unidades singulares, parece que es algo más complejo que eso: alguien sin conocimientos de física no podría comprender física cuántica, por ejemplo. Lo mismo se aplica para otros conocimientos. La religión cristiana, por ejemplo, se relaciona con otros conceptos como la piedad, el perdón, la compasión, etc. Es decir que, si hipersimplificamos y entendemos a la cultura sólo como unidades que se replican de manera egoísta, parece que dejamos de lado un aspecto importante de ésta, que es, precisamente, el modo en el que los elementos culturales se transforman y relacionan. La memética, ya lo hemos mencionado, es imposible de dar cuenta del linaje de los memes y tampoco nos dice cómo se relacionan, se limita a señalar que son replicantes egoístas.

El débil acuerdo de sus partidarios respecto de algunos puntos básicos, así como el apoyo prácticamente nulo de la ciencia empírica que nos permita optar por una postura u otra (considerando que se trata de establecer una ciencia de los memes), hacen difícil tomar a la memética como a un programa de investigación prometedor, por lo menos hasta que se aclaren, con explicaciones sólidas, algunas dudas acerca de las unidades de transmisión cultural, como las mencionadas.

## **Conclusiones**

En este trabajo nos hemos esforzado por mostrar cómo las metáforas que Dennett utiliza para pensar a la conciencia y al ego representan el paso inicial de una explicación filosófica que busca apoyarse en la ciencia empírica, una vez analizados sus presupuestos metafísicos, explicación que intenta desplazar a las visiones tradicionales acerca de tales fenómenos.

Tales metáforas van del modelo del pandemonium hasta la teoría del eco de la conciencia, mostrando algunas carencias, como en el caso de explicar la memoria o la comunicación entre regiones cerebrales. Además, parece que tales imágenes, incluyendo la distinción –metafórica también– que señala que la conciencia y el yo son una especie de *software* corriendo en el cerebro, que funge como *hardware*, podrían ser abandonadas ante los avances en la ciencia empírica que tiene Dennett en mente: la psicología cognitiva sub-personal, o los avances respecto de la explicación de la memoria, tema que, como vimos, nuestro autor nunca aborda detalladamente.

Vimos también cómo el profesor de Tufts trata de analizar y de desechar los presupuestos metafísicos –tanto de la ciencia como del sentido común– que se relacionan normalmente con la conciencia y el ego, cuyo error principal es, de acuerdo con el profesor de Tufts, suponer que hay una línea divisoria o *cadena* causal que divide lo consciente de lo no-consciente, cuando parece ser que se trata, más bien, de una *red* causal, con múltiples vías, procesos paralelos continuos, integración y deterioro de contenidos de la percepción, etc., y que por lo tanto no puede haber tal división.

Si tomamos en serio a la argumentación dennettiana, la conciencia y el yo no son, de ningún modo, un punto central en el cerebro, una especie de espacio donde reside la conciencia humana (como intenta mostrar con los experimentos de

Libet), mucho menos una sustancia pensante, sino un producto de la evolución: son producto del cerebro, pero no origen de las percepciones.

El ego y la conciencia humanos son configurados primordialmente desde la cultura y, aunque la conciencia encuentra su criterio clave en la memoria, es con ayuda del lenguaje y de la cultura que se potencia su desarrollo, dando paso al ego humano, una narración urdida desde fuera, construida socialmente, que funge como un mecanismo de autodefinition y autocontrol. Pero lo anterior no implica una dicotomía a la base del proceso evolutivo, porque tanto el ego como la conciencia se dan en grados en la naturaleza y surgen del control y la definición que vemos en todos los organismos.

Encontramos en Dennett una postura que pretende eliminar la antropomorfización que ve en la conciencia y el ego humanos un salto cualitativo en relación a otras especies, introduciendo así una especie de gradualismo, que el americano considera como una de sus más importantes aportaciones a la filosofía. Pero es también allí, en el darwinismo universal de Dennett, donde encontramos una falla decisiva, pues su postura, extrapolada, lo lleva a sostener una serie de metáforas que él toma literalmente: hablamos de la memética.

Aunque el profesor de Tufts sostiene que los memes no son una metáfora, sino una verdad literal acerca de la evolución, creemos que pudo verse en este trabajo que, mientras sus partidarios no encuentren una manera de dar inicio a una ciencia de los memes, o que al menos se determine qué es una unidad de transmisión cultural y cómo se trasmite y replica, la postura que afirma que los memes son elementos análogos a los genes difícilmente será tomada como una vía estimable para explicar la evolución cultural.

Otra parte del darwinismo de Dennett es su adaptacionismo, cuyas afirmaciones y confianza en la ciencia futura tratamos de matizar. Aunque el adaptacionismo es apoyado por la Actitud Intencional, que es una herramienta útil para predecir y

explicar a otros humanos, a otras especies e incluso algunos rasgos del mecanismo evolutivo, parece un recurso demasiado cómodo y simple para explicar fenómenos tan complejos como el surgimiento de la conciencia refleja, por ejemplo.

A pesar de que a nuestro autor le parece una buena idea incluir en los intentos de explicación evolucionista rasgos pluralistas (interdisciplinarios) como la exaptación y la deriva genética, nunca lo hace. Esto es, si bien el gradualismo que Dennett pretende introducir para pensar a la conciencia y al yo es, según nosotros, valioso en cuanto a que pone en duda a la tradición que ve en tales fenómenos unas diferencias cualitativas respecto de otras especies, parece demasiado simple en algunas de sus explicaciones. El punto sería, pues, buscar explicaciones interdisciplinarias y evitar la monocausalidad de tal gradualismo. En consecuencia, el evolucionismo del americano define lo que considero algunas de sus virtudes, aunque también algunos de sus puntos más débiles.

Pero ¿qué puede haber más importante en la postura de Dennett, o la de cualquier otro filósofo, que señalar cuál es la función primordial de su propia disciplina?

En el caso del autor de *Consciousness Explained*:

1. Las metáforas que propone se van modificando y, en última instancia, pareciera que pudiéramos, en un momento dado, prescindir de muchas de ellas con el avance de la ciencia empírica: este es el optimista legado que quisiéramos atender del profesor de Tufts.
2. Nuestro análisis nos llevó a matizar los logros del americano, en particular respecto de su adaptacionismo, lo que nos invitó a considerar con cautela las hipótesis que la Actitud Intencional apoya respecto de aquél.

3. Por último, criticamos el excesivo aprecio por la memética propuesta originalmente por Dawkins, así como el uso que de esta teoría hace Dennett como parte importante de su darwinismo universal. Creemos que éste puede seguir siendo una hipótesis fructífera sin la memética.

Por lo anterior, concluimos positivamente proponiendo que se puede rescatar el proyecto inicial de nuestro autor, proyecto consecuente con lo que él cree que ha de ser la función primordial de la filosofía: aquella que, parafraseando a Locke, labora despejando –de manera acumulativa, aunque modesta, añadiríamos– el sendero por el que la ciencia empírica contribuye a la explicación de muchos fenómenos. En el caso particular de la conciencia y del ego, liberando a las pesquisas *ad hoc* de las obstrucciones producidas por algunos lastres filosóficos, –considerando que la empresa que Dennett se propone es eliminar el modelo del Teatro Cartesiano– como la idea de que alguno, o ambos de aquéllos, sean un punto central en el cerebro, así como la posición que afirma que los fenómenos que manifiestan a la conciencia o al ego son únicamente humanos.

## **Bibliografía**

- Aunger, Robert (2002). *The Electric Meme. A new Theory of How We Think*. The Free Press, a division of Simon & Schuster.
- Blackmore, Susan (2005). *Conversations on Consciousness*. Oxford University Press.
- Damasio, Antonio. *Descartes' error. Emotion, Reason and the Human Brain*. New York: Grosset/Putnam,1994.
- \_\_\_\_\_ . - *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. First Harvest Edition, 2000.
- Dawkins, Richard (1976). *The Selfish Gene*. Oxford University Press.
- \_\_\_\_\_ . - (1982). *The Extended Phenotype*. Oxford University Press.
- \_\_\_\_\_ . – (2003). *A Devil's Chaplain*. Mariner Books edition. Houghton-Mifflin Company.
- Delius, Juan (1991). *The Nature of Culture*. En *The Tinbergen Legacy*. Chapman & Hall. Londres.
- Dennett, Daniel (1981). *Brainstorms. Philosophical Essays on Mind and Psychology*. First MIT Press Edition.
- \_\_\_\_\_ . - (1984). *Elbow Room. The Varieties of Free Will Worth Wanting*. Oxford University Press.
- \_\_\_\_\_ . - (1989). *The Intentional Stance*. Bradford Books. First MIT Press paperback edition, Cambridge, Massachusetts.
- \_\_\_\_\_ . - (1991). *Consciousness Explained*. Back Bay Books/Little, Brown and Company, New York.
- \_\_\_\_\_ . - (1995). *Darwin's Dangerous Idea. Evolution and the meanings of life*. Penguin Books.
- \_\_\_\_\_ . - (1996). *Kinds of Minds. Towards an understanding of consciousness*. BasicBooks.

- \_\_\_\_\_ . - (1998). *Brainchildren: essays on designing minds*. Bradford Books.
- \_\_\_\_\_ . - (2005). *Sweet Dreams. Philosophical Obstacles to a Science of Consciousness*. Bradford Books. The MIT Press. Cambridge, Massachusetts.
- Freeman, Walter J. (2000). *How brains make up their minds*. Columbia University Press.
- Gazzaniga, Michael. *The Mind's Past*. University of California Press, 1998.
- \_\_\_\_\_ . - (2000). *The New Cognitive Neurosciences*. The MIT Press.
- \_\_\_\_\_ . - (2008). *Human. The Science behind what makes us unique*. Harper-Collins e-Books,
- McCarthy, Joan. (2007). *Dennett and Ricoeur on the Narrative Self*. Humanity Books.
- Richerson, Peter J. and Boyd, Robert (2005). *Not By Genes Alone. How Culture Transformed Human Evolution*. University of Chicago Press.

### Recursos electrónicos

- Dennett, Daniel (1989b). *The Origins of Selves*, *Cogito*, 3, 163-73.  
<http://ase.tufts.edu/cogstud/papers/originss.htm>
- \_\_\_\_\_ . - (1992). *The Self as the Center of Narrative Gravity*, in *Self and Consciousness: Multiple Perspectives*, Kessel, Cole, Johnson eds., Erlbaum.  
<http://ase.tufts.edu/cogstud/papers/selfctr.htm>
- \_\_\_\_\_ . - (2003). *The Self as a Responding—and Responsible—Artifact*. *Annals New York Academy of Sciences* 1001: 39-50  
<http://ase.tufts.edu/cogstud/papers/SelfasaResponding.pdf>
- \_\_\_\_\_ . - (2008). Entrevista a Daniel Dennett.  
<http://www.youtube.com/watch?v=9O1qZYKSQmc>
- \_\_\_\_\_ . - (2009). *The Part of Cognitive Science that is Philosophy*

[http://ase.tufts.edu/cogstud/papers/tops\\_1015.pdf](http://ase.tufts.edu/cogstud/papers/tops_1015.pdf)

- \_\_\_\_\_ . - (2009b). *Intentional Systems Theory*, Oxford Handbook of the Philosophy of Mind B. McLaughlin, A. Beckermann, S. Walter, eds., OUP, , pp. 339-50  
<http://ase.tufts.edu/cogstud/papers/intentionalsystems.pdf>
- Gould, Stephen Jay; Lewontin, Richard (1979). *The Spandrels of San Marco and the Panglossian Paradigm: A Critique Of The Adaptationist Programme*, Proceedings Of The Royal Society of London, Series B, Vol. 205, No. 1161 Pp. 581-598.  
<http://ethomas.web.wesleyan.edu/wescourses/2004s/ees227/01/spandrels.html>
- Lewontin, Richard (1978). *Fitness, Survival, and Optimality* in D. H. Horn, R. Mitchell, and G. R. Stairs, eds., *Analysis of Ecological Systems*. Cincinnati: Ohio State University Press.  
[http://www.kli.ac.at/theorylab/EditedVol/H/HornDJ\\_78.html](http://www.kli.ac.at/theorylab/EditedVol/H/HornDJ_78.html)
- Libet, Benjamin (1999). *Do we Have Free Will?* Journal of Consciousness Studies, 6, No. 8–9.  
[www.imprint-academic.com/jcs](http://www.imprint-academic.com/jcs)
- Murillo, José Ignacio; Jiménez-Amaya, José Manuel (2008). *Tiempo, conciencia y libertad: consideraciones en torno a los experimentos de Libet y colaboradores*. Acta Philosophica. Rivista internazionale di filosofia fascicolo II, volumen 17. Pontificia Universitat Della Santa Croce.  
[http://www.actaphilosophica.it/en/fascicoli/08/08\\_2\\_murillo.php](http://www.actaphilosophica.it/en/fascicoli/08/08_2_murillo.php);  
<http://www.unav.es/cryf/libet.html#4>
- Soler Gil, Francisco José (2009). *Relevancia de los experimentos de Benjamin Libet y de John-Dylan Haynes para el debate en torno a la libertad humana en los procesos de desición*. Revista de Filosofía Thémata, número 41.  
<http://www.institucional.us.es/revistas/revistas/themata/pdf/41/34soler.pdf>



**Universidad Autónoma Metropolitana  
Unidad Iztapalapa  
Posgrado en Humanidades  
Línea Historia y Filosofía de la Ciencia**

***El evolucionismo de Dennett: las teorías del ego y de la conciencia***

**Tesis de Maestría  
(Idónea comunicación de resultados)**

**Que presenta  
Lic. Víctor Manuel Romero Sánchez**

**Director  
Dr. Jorge Martínez Contreras**

**Jurado  
Dr. Cuauhtémoc Lara Vargas  
Dr. Jonatan García Campos  
Dr. José Luis Vera Cortés**

**Julio de 2010**