



<https://doi.org/10.15446/ideasyvalores.v75n190.125687>

LA DISCONTINUIDAD DE NIVELES EN CIENCIA COGNITIVA*



THE DISCONTINUITY OF LEVELS IN COGNITIVE SCIENCE

BRYCE S. GESSELL**

Southern Virginia University - Buena Vista - USA

FELIPE DE BRIGARD***

Duke University - Durham - USA

Cómo citar este artículo:

MLA: Gessell, Bryce, y Felipe De Brigard. "La discontinuidad de niveles en ciencia cognitiva." *Ideas y Valores*, vol. 75, no. 190, 2026, pp. 115–132.

APA: Gessell, B., y De Brigard, F. (2026). La discontinuidad de niveles en ciencia cognitiva. *Ideas y Valores*, 75(190), 115–132.

discontinuidad de niveles en ciencia cognitiva. *Ideas y valores*, 190.

CHICAGO: Gessell, Bryce, y Felipe De Brigard. 2026. "La discontinuidad de niveles en ciencia cognitiva." *Ideas y Valores* 75 (190): 115–132.

* Este artículo apareció originalmente en la revista *Teorema*, en un número especial sobre el libro *From Bacteria to Bach and Back: The evolution of Minds*. Su referencia original es: Gessell, Bryce, y Felipe de Brigard. "The Discontinuity of Levels in Cognitive Science." *Teorema* 373 (2018): 151–165. La presente traducción aparece acá por primera vez gracias al generoso permiso de la revista *Teorema*.

** bsg16@duke.edu / ORCID: <https://orcid.org/0000-0003-4424-5627>.

*** felipe.debrigard@duke.edu / ORCID: <https://orcid.org/0000-0003-0169-1360>.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

RESUMEN

Comenzamos caracterizando la concepción “funcionalista homuncular” de la mente de Dennett, tal como es descrita en su obra temprana. A continuación, la comparamos con la bosquejada en *From Bacteria to Bach and Back*. Argumentamos que los cambios recientes en la posición de Dennett han producido cierta tensión en su visión de las descomposiciones funcionales. Supuestamente, las descomposiciones funcionales basadas en la postura intencional terminan en un nivel inferior, no inteligente, que puede explicarse mecánicamente. Sin embargo, puesto que Dennett cree ahora que las neuronas podrían tener que ser descritas en términos intencionales, no queda claro si nuestras explicaciones de las funciones cognitivas van a casar con nuestras explicaciones de la conducta de neuronas y redes. Exploramos las consecuencias de esta tensión para la teoría de Dennett y para la neurociencia cognitiva en general.

Palabras clave: neurociencia cognitiva, funcionalismo, ontología cognitiva, estrategia intencional, psicología.

ABSTRACT

We begin by characterizing Dennett’s “homuncular functionalist” view of the mind, as described in his early work. We then contrast that view with the one outlined in *From Bacteria to Bach and Back*. We argue that recent changes in Dennett’s view have produced tension in the way he conceives of functional decompositions. Functional decompositions based on the intentional stance are supposed to reach a bottom, “dumb” level which can be explained mechanically; however, since Dennett now believes that neurons may need to be described intentionally, it is not clear whether our explanations of cognitive functions can ever align with our explanations of neuronal and network behaviors. We explore the consequences of this tension for Dennett’s view, and for cognitive neuroscience in general.

Keywords: Cognitive Neuroscience; Functionalism; Cognitive Ontology; Intentional Stance; Psychology.

Introducción

La trayectoria de Dennett abarca ya varias generaciones; sorprendentemente, el próximo año se cumplirán *cincuenta* años de la publicación de su primer libro, *Contenido y Conciencia* (Dennett, 1969). En esa obra y en muchas posteriores, Dennett abordó una enorme variedad de temas, desde cuestiones tradicionales de la filosofía de la mente hasta problemas en la intersección de otras disciplinas, como la psicología, la neurociencia, la ciencia cognitiva, la biología evolutiva y más. Al ser dos de sus antiguos alumnos, podemos dar fe de que su enseñanza y mentoría abarcan el mismo amplio espectro.

From Bacteria to Bach and Back (FBBB) reúne numerosos hilos del pensamiento de Dennett, tejiendo una historia sobre el origen de la mente y la conciencia humana a través de los memes que permean a individuos y culturas. Si esta historia les resulta familiar, es porque Dennett ha estado contando una versión de ella desde el principio:

He dedicado medio siglo, toda mi vida académica, a este proyecto en una docena de libros y cientos de artículos que abordan diversas piezas del rompecabezas, sin lograr convencer a muchos lectores a pasar de un cauteloso agnosticismo a una calmada convicción. Sin desanimarme, lo intento una vez más y esta vez voy a contar la historia completa. (Dennett 2017 16)

Pero al contar “la historia completa”, FBBB trasciende la obra anterior de Dennett y transforma sus perspectivas al integrarlas plenamente con las investigaciones recientes de las ciencias.

En algunos puntos, sin embargo, ciertas afirmaciones sobre “la historia completa” presentadas en FBBB difieren de las afirmaciones y opiniones que Dennett sostuvo en el pasado. En la mayoría de los casos, estas modificaciones son inocuas, pero al menos una de ellas podría tener consecuencias más graves de lo que el propio Dennett parece reconocer. Nos referimos a su revisión del “funcionalismo homuncular” en el capítulo 8. En este breve comentario, sugerimos que el cambio de opinión de Dennett sobre el funcionalismo homuncular pone de manifiesto una tensión interna entre las estrategias intencional y de diseño, que incluso podría cuestionar la legitimidad de la neurociencia cognitiva —al menos tal como se practica actualmente— como una estrategia fiable para desentrañar los mecanismos neuronales de la mente.

Con este fin, comenzamos —en la sección II— con una revisión del funcionalismo homuncular de Dennett, tal como se entendía antes de FBBB. A continuación, en la sección III, reconstruimos cuidadosamente su reexamen del funcionalismo homuncular en FBBB, destacando las diferencias entre las formulaciones original y la nueva. Posteriormente, en la sección IV, argumentamos que la nueva formulación sugiere

firmemente que la visión jerárquica resultante de una descomposición funcional descendente —es decir, de las personas a las neuronas— puede no coincidir con la visión jerárquica que emerge de una descomposición ascendente —es decir, de las neuronas a las personas—. Las consecuencias de esta discrepancia —como argumentamos en la sección V— son profundas, ya que no solo ponen en tela de juicio si se puede superar la brecha entre la estrategia intencional y la de diseño, sino, más generalmente, si alguna vez se podrá superar el desafío científico de mapear predicados mentales a categorías neuronales.

El funcionalismo de Dennett

Las ideas de Daniel Dennett no encajan fácilmente dentro de los estrictos límites de las etiquetas filosóficas. ¿Es realista o antirrealista acerca de las actitudes proposicionales? ¿Es instrumentalista o eliminativista respecto de la mente? ¿Es realista o antirrealista acerca de la conciencia? Sin embargo, dado que las antologías deben publicarse y los capítulos deben encajar en secciones, no sorprende que las inclinaciones de Dennett hacia explicaciones computacionales de los procesos mentales hicieran que sus primeros escritos parecieran pertenecer junto a los de los primeros funcionalistas, como Hilary Putnam (*"Minds and Machines"*; *"Psychological Predicates"*), David Armstrong y también David Lewis (*"Psychophysical and Theoretical Identifications"*). No obstante, desde el principio hubo diferencias fundamentales entre la versión del funcionalismo de Dennett y la de lo que podríamos llamar los funcionalistas de tabla de máquina. Putnam (*"Minds and Machines"*; *"Psychological Predicates"*), por ejemplo, siguiendo la idea de Alan Turing de que la inteligencia podía entenderse como procesos algorítmicos complejos, propuso que los estados mentales podían corresponder a estados de una tabla de máquina de Turing. Posturas relacionadas fueron defendidas posteriormente por Armstrong y también por Lewis (*"How to Define Theoretical Terms"*, *"Psychophysical and Theoretical Identifications"*). La teoría causal de la mente de Armstrong, por ejemplo, sostenía que un estado mental debía definirse como un "estado que es apto para ser la causa de ciertos efectos o apto para ser el efecto de ciertas causas", independientemente de su realización material.

En este sentido, la teoría causal de Armstrong es similar al funcionalismo de tabla de máquina de Putnam, pues ambos definen los estados mentales en términos causales o computacionales, independientemente del material en el que se realicen. De manera importante, la idea de que la misma tabla de máquina pudiera ser implementada por distintos realizadores —sean neuronas, chips o tubos de vacío— motivó a los funcionalistas a postular dos tipos de identificaciones mentales. Primero, estaban las identificaciones psicofísicas: relaciones

uno a uno entre estados mentales, entendidos como estados definidos computacionalmente (es decir, causalmente) de una tabla de máquina de Turing, y los realizadores particulares que los producen. Aquí, las ideas de Lewis resultaron cruciales. En términos generales, su sugerencia es que, para identificar los estados mentales con sus realizadores particulares, primero hay que

reunir todas las trivialidades [platitudes] que puedas imaginar acerca de las relaciones causales entre estados mentales, estímulos sensoriales y respuestas motoras, [y luego] añadir también las trivialidades según las cuales un estado mental cae bajo otro —‘el dolor de muelas es un tipo de dolor’, y cosas por el estilo. (Lewis 1972 256)

Estas trivialidades de la psicología popular constituyen los términos teóricos cuyas definiciones se dan en términos de roles causales particulares. Finalmente, para identificar el realizador físico de un término teórico particular dentro del conjunto de trivialidades psicológicas populares, y así establecer una identificación psicofísica entre ambos, era necesario obtener la oración de Ramsey que incluyera el término psicológico relevante, y luego identificar la entidad designada por su variable existencial acotada. La entidad designada por el cuantificador existencial acotado sería el correlato ontológico del estado mental referido por el término psicológico correspondiente. Así, si un término, *T*, pertenece a la lista de trivialidades psicológicas, entonces el realizador físico de *T* sería la entidad designada por la variable ligada, *x*, de la oración de Ramsey, $\exists!xP(x)$.

La segunda identificación que el funcionalismo de tabla de máquina defiende es la identidad psicológica, esto es, relaciones uno a uno entre estados mentales. Específicamente, los funcionalistas de tabla de máquina estaban comprometidos con la afirmación de que dos organismos, independientemente de su constitución física, podrían, sin embargo, estar en el mismo estado mental siempre y cuando se encontraran en el mismo estado de una tabla de máquina idéntica. Así como dos computadores hechos de distintos materiales podrían, no obstante, hallarse en el mismo estado computacional, dos organismos completamente distintos podrían estar en el mismo estado de dolor o mantener la misma creencia mientras los estados de sus tablas de máquina fueran computacionalmente idénticos. Además, la teoría parece comprometida con la afirmación de que dos entidades, sin importar su composición física, podrían en principio instanciar exactamente la misma vida mental siempre y cuando ambas instancien exactamente la misma tabla de máquina. Asimismo, dos personas estarían en el mismo estado mental si y solo si se hallen en el mismo estado de máquina. Esto implica que mentes como las nuestras deben tener una y

solo una única descripción funcional. Entidades con tablas de máquina de Turing que no correspondan a las nuestras no contarían como poseedoras de mentes como las nuestras.

Años después, Ned Block señaló acertadamente que esta visión, que se suponía iba a reducir el liberalismo del conductismo, termina siendo también demasiado liberal, pues atribuye erróneamente mentalidad a entidades que aparentemente carecen de ella. Pero la alternativa —restringir las tablas de máquina únicamente a aquellas entidades con psicologías isomorfas a las nuestras— resulta ser tan chovinista como la teoría de la identidad, pues terminamos negando mentalidad a entidades que aparentemente la poseen.

Sin embargo, las objeciones de Block al funcionalismo de máquina nunca afectaron realmente la postura computacionalista del propio Daniel Dennett, ya que él no cree que las identidades psicológicas —es decir, el mapeo uno a uno entre los estados mentales de distintos organismos— deban encontrarse al nivel de la tabla de máquina, sino al nivel de la estrategia intencional. Así, dos entidades podrían estar en estados de máquina completamente diferentes y, por tanto, requerir descripciones funcionales totalmente distintas desde la postura del diseño, y aun así podrían compartir las mismas descripciones intencionales desde la estrategia intencional (Dennett 2005). Para Dennett, por ende, las identificaciones psicológicas no dependen de que ejecutemos exactamente el mismo algoritmo mental; provienen de descripciones intencionales uniformes generadas desde la estrategia intencional (De Brigard 2007 132).

Pero entonces, si dos organismos pueden estar en exactamente el mismo estado intencional mientras difieren en sus estados funcionales subyacentes, ¿cómo podemos mapear alguna vez los términos mentales con los físicos? El enfoque de Dennett para responder esta pregunta encuentra inspiración en la manera en que opera la inteligencia artificial:

El programador de IA empieza con un problema caracterizado intencionalmente, y por ello ve sin tapujos al computador antropomórficamente: si *resuelve* el problema dirá que ha diseñado un computador que puede entender preguntas en inglés. Su primer y más alto nivel de diseño descompone el computador en subsistemas, cada uno de los cuales recibe tareas caracterizadas intencionalmente; compone un diagrama de flujo de evaluadores, recordadores, discriminadores, supervisores y similares. Estos son *homúnculos* con ganas; el diseño de nivel más alto descompone el computador en un comité o ejército de homúnculos inteligentes con propósitos, información y estrategias. Cada homúnculo, a su vez, se analiza en homúnculos *más pequeños*, pero, más importante aún, en homúnculos menos astutos. Cuando se alcanza el nivel en el que los homúnculos no son más que sumadores y restadores, cuando solo necesitan la intelligen-

cia para escoger el mayor de dos números cuando se les indica, entonces han sido reducidos a funcionarios ‘que pueden ser reemplazados por una máquina’. La ayuda para la comprensión que brinda antropomorfizar los elementos prácticamente se disipa en este punto, y una visión mecanicista de los procesos se vuelve manejable y comprensible. [Dennett 1978 80-81]

Lycan llama a esta estrategia *funcionalismo homuncular* (por ejemplo, “Form, function, and feel”; *Consciousness*), ya que sugiere que un sistema intencional, como nosotros, puede subdividirse jerárquicamente en subsistemas anidados, donde cada nivel inferior requiere menos vocabulario intencional para su descripción que el nivel anterior, hasta que la jerarquía llega al nivel de una descripción que no requiere vocabulario intencional alguno: el nivel de las neuronas.

De manera importante, el funcionalismo homuncular no es solo un planteamiento sobre la estructura de nuestra mente: también es una postura que guía nuestras prácticas científicas en psicología cognitiva y neurociencia. Como nos recuerda Dennett:

La tarea de la psicología es explicar la percepción humana, el aprendizaje, la cognición, y así sucesivamente, en términos que finalmente unan la teoría psicológica con la fisiología de un modo u otro, y existen dos estrategias generales que uno podría adoptar: una estrategia *de abajo hacia arriba* [*bottom-up*], que empieza con alguna unidad básica y bien definida o átomo teórico para la psicología, y construye estos átomos en moléculas y agregados más grandes que puedan dar cuenta de los fenómenos complejos que todos observamos; o una estrategia *de arriba hacia abajo* [*top-down*], que comienza con una descomposición más abstracta de los niveles más altos de organización psicológica, y espera analizar estos en sistemas o procesos progresivamente más pequeños y detallados hasta llegar finalmente a elementos familiares para los biólogos. (Dennett 1978 110)

Junto con Simon y también Marr, y anticipándose a Shallice, además de a Cummins, Dennett aboga por un enfoque de descomposición funcional de arriba hacia abajo que refleja la estructura del funcionalismo homuncular. Y sostiene que “la estrategia de abajo hacia arriba en psicología es poco probable que resulte muy fructífera”, porque los dos “intentos mejor desarrollados [...] son ahora ampliamente considerados como estancados” (Dennett 1978 110). El primer intento fue el conductismo estímulo–respuesta, que, según Dennett, no tendrá éxito porque las asociaciones estímulo–respuesta resultan ser el tipo equivocado de átomos psicológicos básicos. El segundo intento —más relevante para los propósitos actuales— es lo que él llamaba entonces “psicología fisiológica de señales neuronales”, y pensaba que

aunque las sinapsis y los trenes de impulsos sean átomos perfectamente adecuados, simplemente hay demasiados de ellos y sus interacciones son demasiado complejas para estudiarlas una vez se abandonan las periferias aferentes y eferentes y se intenta dar sentido al centro crucial. (Dennett 1978 110)

En resumen: la perspectiva funcionalista a la que llegamos difiere del funcionalismo basado en tablas de máquinas en aspectos fundamentales. Primero, rechaza las identificaciones psicológicas porque la uniformidad de nuestras descripciones intencionales no depende de compartir la misma tabla de máquinas. Dos sistemas intencionales pueden diferir en sus descripciones físicas y computacionales y, sin embargo, instanciar los mismos estados mentales. La segunda diferencia —que se deriva de la primera— es que las identificaciones psicofísicas no se desarrollarán como sugirió Lewis, ya que, incluso si pudiéramos recopilar todos los lugares comunes de la psicología popular, es improbable que existan oraciones de Ramsey equivalentes entre individuos, de modo que se pueda identificar un único realizador físico mediante una variable ligada. En consecuencia, Dennett aboga por una estrategia empírica, en lugar de una *a priori*, para identificar los realizadores físicos de los estados intencionales: una descomposición funcional descendente [*top-down*] que refleje la estructura de su funcionalismo homuncular.

Funcionalismo homuncular revisado

Desde la década de 1970, la neurociencia se ha desarrollado a un ritmo asombroso, y ahora sabemos mucho más sobre la neurobiología, la neuroquímica y la neuroanatomía funcional de los cerebros humanos y no humanos que nunca antes. No sorprende que Daniel Dennett haya estado muy al tanto de estos desarrollos, lo cual quizá explique por qué, cuarenta años después de la publicación de *Brainstorms* y de la introducción de su funcionalismo homuncular, decidió que era momento de revisarlo en FBBB:

Todavía creo que esto va en la dirección correcta, pero he llegado a lamentar —y rechazar— algunas de las connotaciones de los dos términos que utilicé: ‘comité’ y ‘máquina’. La burocracia cooperativa que sugiere el primero, con sus claras relaciones jerárquicas (una imagen reforzada por los diagramas de flujo sin concesiones de los modelos de la ciencia cognitiva clásica) capturaba el sueño de la GOFAT [Good Old Fashion Artificial Intelligence] de arriba hacia abajo [*top-down*], pero sugería un tipo de eficiencia profundamente no biológica. La extraña inversión del razonamiento de Turing sigue intacta: eventualmente, en nuestra cascada de descomposición, llegamos a elementos cuyas tareas son tan rígidas y

rutinarias que ‘pueden ser reemplazados por una máquina’, tal como el diligente computador humano de Turing. Las partes móviles más simples dentro de las neuronas —las proteínas motoras, los microtúbulos y similares— sí son autómatas sin motivación, como las escobas marchantes en *El aprendiz de brujo*, pero las propias neuronas, en sus miles de millones, cumplen roles más emprendedores e idiosincráticos que los obedientes empleados que yo imaginaba, y ese hecho tiene implicaciones importantes para la arquitectura computacional de los cerebros. (Dennett 2017 162)

¡Y, sin duda, las tiene! Las neuronas son mucho más complejas, y su comportamiento es mucho más agencial —aunque inconsciente— de lo que creíamos hace cuarenta años. Ahora sabemos que los modelos neuronales simples que heredamos de McCulloch y Pitts son simplemente erróneos, pues la dinámica neuronal es compleja, y los modelos paramétricos simples parecen incapaces de capturar dicha complejidad.

Consideremos algunos casos de esta complejidad. Según la visión de las neuronas como simples interruptores, estas células pueden decir “sí” o “no” a los estímulos entrantes al disparar o no disparar, respectivamente. Una visión un poco más elaborada toma a las neuronas como compuertas lógicas que participan, al disparar, en operaciones lógicas como “y”, “o” y “si-entonces”. La visión resultante concibe el sistema nervioso como una especie de tarjeta madre en una computadora, cuyas operaciones están cableadas en las células y en sus conexiones. Las contingencias biológicas del entorno celular se consideran intrascendentes y pueden ignorarse.

Sin embargo, la evidencia científica acumulada en las últimas décadas sugiere con contundencia que el comportamiento de las neuronas es mucho más matizado de lo que se creía, y que estos matices pueden desempeñar roles importantes en la cognición. Las neuronas, por ejemplo, no solo están conectadas sinápticamente a otras neuronas, sino también mediante enlaces estructurales llamados uniones de brechas (*gap junctions*) (Shimizu y Stopfer). Las señales pueden pasar de una célula a otra a través de estos enlaces, pero, lo que es importante, estas señales no son tan cuantificables ni tan regulares como los potenciales de acción. Además, las neuronas pueden liberar moléculas neurotransmisoras en espacios celulares para influir en el comportamiento de circuitos o redes completas. Este comportamiento neuronal, llamado “transmisión por volumen”, se opone a la visión de las neuronas como simples interruptores o compuertas con objetivos claros, que puedan modelarse en un nivel altamente abstracto. También se ha descubierto recientemente que las neuronas pueden compartir señales mediante el gen *Arc* y sus proteínas (Pastuzyn *et al.*). Este medio de comunicación puede permitir que neuronas que no están activas sufran

cambios como si lo estuvieran, al producir proteínas en respuesta a recibir arnm de otras neuronas. Por último, es probable que los mecanismos neuronales no se comporten de manera tan determinista como los modelos simples sugerían (Gessell), y las neuronas pueden hacer *multiplexing*, pues parecen transportar señales de más de un estímulo a la vez (Caruso *et al.*).

No hace falta decir que esta brevísima revisión apenas toca la superficie de lo que los hallazgos neurobiológicos y neuroquímicos recientes sugieren de forma abrumadora: que las neuronas no son simples interruptores de encendido y apagado. A partir de esta nueva comprensión de las neuronas, se siguen dos consecuencias muy importantes que afectan la formulación original del funcionalismo homuncular de Daniel Dennett.

Primero, debemos reconsiderar el alcance real de la realizabilidad múltiple. Cuando el funcionalismo de tabla de máquina estaba de moda, los filósofos afirmaban que era concebible tener mentes implementadas en todo tipo de sustratos distintos, incluidos los mercados financieros de Bolivia o los habitantes de China (Block). Pero, basándose en los desarrollos recientes en neurobiología y neuroquímica, Dennett ha insistido —al menos desde 2003— en lo que él llama funcionalismo minimalista: la idea de que, aunque la realizabilidad múltiple, tal como fue concebida inicialmente por los funcionalistas, puede ser verdadera en principio, en realidad es probable que ningún otro material pueda hacer lo que las neuronas hacen (véase, también, Clark; Godfrey-Smith).

Eso no significa, como él señala en la cita anterior, que la “extraña inversión del razonamiento” de Turing fuera equivocada: sigue intacta. Eventualmente, nuestra descomposición funcional hallará “elementos cuyas tareas son tan rígidas y rutinarias que ‘pueden ser reemplazados por una máquina’” (Dennett 2017 162). Pero estos elementos se encuentran dentro de las neuronas, a nivel de proteínas, microtúbulos y similares, cuyo comportamiento —esperemos— será finalmente capturado por descripciones algorítmicas formuladas adecuadamente desde la postura del diseño.

Si esto es así, entonces una segunda consecuencia se sigue para el funcionalismo homuncular, a saber: que, contrariamente a su formulación inicial, el comportamiento complejo de una neurona puede capturarse mejor desde la estrategia intencional y no desde la estrategia del diseño. Para enfatizar este punto, Dennett alude amigablemente, por ejemplo, a la noción de Fitch de “nano-intencionalidad” y a las ideas de Seung sobre “neuronas egoístas” y “sinapsis hedonistas”, que parecen capturar el ahora evidente hecho de que las neuronas, por así decirlo, tienen mente propia.

Jerarquías dispares

Superficialmente, estas dos consecuencias no parecerían demasiado revolucionarias para la noción original del funcionalismo homuncular. Después de todo, ¿no está Dennett simplemente moviendo un poco los límites al decirnos que no son las neuronas, sino las cosas dentro de las neuronas, las que constituyen el nivel más básico de la descomposición funcional? En cierto sentido, sí; sin duda está enfatizando que la estructura jerárquica asumida por las arquitecturas GOFAI es excesivamente simplista y antinatural:

El diseño inteligente de arriba hacia abajo [top-down] de la computadora clásica creó una maravilla hipercompetente pero profundamente no biológica. Esto no se debe a que las computadoras estén hechas del tipo equivocado de material, sino a que el material del que están hechas está organizado en el tipo equivocado de jerarquías: el tipo de burocracias planificadas que pueden ser “máquinas bien engrasadas” pero que dependen de la regimentación de las partes móviles, suprimiendo tanto la exploración como la improvisación en todos los niveles. (Dennett 2017 163)

Obsérvese que la imagen de descomposición funcional ofrecida aquí sugiere que la jerarquía de niveles puede verse muy diferente cuando se observa desde arriba en comparación con cuando se observa desde abajo. En su formulación original, el funcionalismo homuncular planteaba una descomposición de procesos cognitivos complejos en subprocesos cada vez más simples, cada uno —literalmente— *más tonto que el anterior*, donde “*más tonto*” se entendía como menos necesitado de descripciones intencionales que el sistema superior. Para cuando llegábamos al nivel microscópico de las neuronas, esperábamos estar en un punto en el que ya no se requirieran descripciones intencionales para captar su comportamiento.

Sin embargo, la historia narrada en FBBB es diferente. Ahora el nivel neuronal no es tan tonto como pensábamos —de hecho, puede que necesitemos la estrategia intencional para captar plenamente la complejidad de los comportamientos de una sola neurona. Si es así, existe una fuerte posibilidad de que la estructura de la jerarquía asumida por el GOFAI simplemente no corresponda a la jerarquía ascendente de un enfoque de abajo hacia arriba, es decir, de las neuronas a las personas. Esto se debe a que nuestras descripciones del comportamiento neuronal serán ellas mismas intencionales. Recordemos que, en *Brainstorms*, la razón para rechazar el enfoque ascendente de la “psicología fisiológica de señales neuronales” era que había demasiadas neuronas con interacciones demasiado complejas. Sin embargo, el mensaje que recibimos fue que, en principio, si pudiéramos estudiar todas las neuronas y sus interacciones, podríamos abrirnos camino hacia arriba, por así decirlo,

y recomponer la jerarquía funcional pasando de los átomos neuronales “tontos”, caracterizados desde la estrategia de diseño, hasta los agentes cognitivos descritos desde la estrategia intencional.

Es posible interpretar la revisión de Dennett en *FBBB* como simplemente sugerir que los verdaderos átomos podrían ser los microtúbulos o las mitocondrias, y que la descomposición jerárquica permanece intacta; simplemente habíamos identificado mal el nivel inferior. Pero nuestra preocupación radica en que, al considerar a las neuronas como agentes, incluso si son skinnerianas (Dennett 2017 165), descubramos que las mejores formas de caracterizar su comportamiento no encajan con ninguno de los subsistemas en los que se suponía que los niveles superiores debían descomponerse desde arriba hacia abajo. En otras palabras: Dennett sugirió que nuestra descomposición funcional de arriba hacia abajo iba a incluir registradores, discriminadores, etc., subsistemas cuyos comportamientos estaban claramente caracterizados como engranajes dentro de un mecanismo mayor y anidado. Pero ahora resulta que, si comenzamos desde el nivel neuronal y avanzamos hacia arriba, podríamos encontrar que las mejores caracterizaciones del comportamiento neuronal no corresponden a ninguno de los mecanismos anidados que se suponía formaban parte del agente como un todo.

Esta posible discrepancia se vuelve más evidente cuando nos preguntamos, con Dennett: “¿qué podría ‘querer’ una neurona?” (Dennett 2017 162). Más generalmente, ¿qué tipo de estados intencionales podrían ayudarnos a explicar su comportamiento complejo? La respuesta de Dennett es que, al igual que sus primos procariotas, las neuronas están diseñadas para lidiar con las *affordances*, o las cosas que les importan. ¿Y qué les importa a las neuronas? Fisiológicamente, las neuronas se preocupan por exactamente aquellas cosas de las que esperaríamos que se preocupara cualquier organismo unicelular: el mantenimiento y la reparación de estructuras celulares, la producción y el uso de energía, etc. Pero las neuronas también se preocupan por la información. En concreto, la sugerencia de Dennett es que las neuronas se preocupan por cablearse entre sí para detectar *affordances*. Las *affordances* en este caso ofrecen información sobre los estímulos que una neurona puede esperar recibir, y por tanto permiten a las neuronas hacer predicciones jerárquicas. Las predicciones anticipan la entrada a las neuronas, y el término “jerárquico” se refiere al hecho de que las predicciones pueden hacerse en diferentes niveles de abstracción. Los bucles de retroalimentación desde áreas más altas de la jerarquía influyen en un código predictivo y lo modifican en respuesta a la precisión de sus predicciones.

Consideremos ahora el enfoque de codificación predictiva a la luz de lo que hemos aprendido sobre el funcionalismo homuncular de Dennett. El comportamiento neuronal es muy complejo, y hemos

llegado a creer que requiere un nivel de descripción más elevado que la simple estrategia de diseño. Es decir, requiere la estrategia intencional. Pero si este es el caso —y si las neuronas están computando predicciones sobre la presencia o ausencia de cosas que les importan— entonces surge la posibilidad de que las categorías funcionales adecuadas para describir el comportamiento de las neuronas, así como de otras estructuras nerviosas ascendentes como áreas cerebrales, redes y sistemas, simplemente no correspondan a los niveles funcionales de descripción asumidos por las descomposiciones de arriba hacia abajo (De Brigard 2012). Estas son las descomposiciones funcionales de la ciencia cognitiva estándar y del GOFAI, que intentamos fundamentar en las actividades de poblaciones neuronales u otras unidades funcionales. Pero lo que a las neuronas les importa, o sobre lo que hacen predicciones, puede no ser relevante para aquello que nos parece importante en las descomposiciones de las funciones cognitivas. Lo que hace el cerebro y cómo está organizado desde abajo hacia arriba puede verse muy diferente de lo que hace y cómo está organizado desde arriba hacia abajo.

Veamos un ejemplo de cómo podría ocurrir esto. Supongamos que nos interesa caracterizar la contribución que una cierta red neuronal hace a la cognición, y el enfoque de codificación predictiva jerárquica sugiere que la función de la red es calcular expectativas para la presencia de cierta propiedad *P*. Desde esta perspectiva de abajo hacia arriba, hemos identificado una unidad funcional —la red— y también la función de la unidad —predecir y comprobar la presencia de *P*—. Si continuáramos identificando unidades funcionales de este modo, podríamos identificar alguna otra red de la que esta, más pequeña, forma parte, de modo que la red mayor tenga también alguna descripción funcional que proporcionaríamos desde abajo hacia arriba.

La pregunta es: ¿por qué debería ser el caso que nuestra descomposición funcional de arriba hacia abajo de un sistema cognitivo mayor —digamos, la memoria— también involucre la propiedad *P*? En otras palabras, ¿y si *P* es algo que tenemos buenas razones para pensar que nuestra red realmente *sí* valora, aunque *P* no aparezca en ninguna descomposición detallada de nuestra capacidad para recordar? ¿Y si las cosas que les importan a las redes neuronales, o incluso a las neuronas individuales, no son las cosas que nos importan a nosotros? El nuevo pensamiento de Dennett sobre el funcionalismo homuncular ha abierto esta posibilidad, porque nuestras descripciones intencionales del comportamiento neuronal deberían ser como descripciones intencionales del comportamiento procariota: solo contienen cosas que le importan al sistema descrito. Así, aunque lo que nos importa a nosotros sin duda aparecerá en las descomposiciones funcionales de nuestros sistemas cognitivos, puede resultar imposible conectar los

subsistemas correspondientes con unidades funcionales neuronales, puesto que esas unidades podrían tener una descripción funcional que les es propia, independiente de nuestro análisis de arriba hacia abajo.

Ontologías cognitivas y el mapeo mente-cerebro

La sección anterior terminó con el problema de alinear descripciones con mecanismos desde dos perspectivas diferentes: una que procede de abajo hacia arriba y otra que procede de arriba hacia abajo. Creemos que este problema puede tener otra consecuencia de gran alcance para la visión de Daniel Dennett.

Como muchos otros interesados en la neurociencia cognitiva, Dennett asume razonablemente que uno de los objetivos principales de esta ciencia es producir un mapeo entre los procesos cognitivos a nivel personal y los mecanismos neuronales a nivel subpersonal. Esto es cierto cuando tratamos los sistemas cognitivos únicamente como sistemas de procesamiento de información, pero también lo es cuando consideramos la fenomenología involucrada. Dennett, por ejemplo, describe la importancia de este mapeo de la siguiente manera:

No tendremos una ciencia completa de la consciencia hasta que podamos alinear nuestras identificaciones, desde la imagen manifiesta, de los estados mentales por sus contenidos con las identificaciones de la imagen científica de las estructuras y eventos subpersonales de información que son causalmente responsables de generar los detalles de la ilusión-de-usuario en la que creemos operar. (Dennett 2017 367)

La suposición del objetivo final de la neurociencia cognitiva es clara aquí: lo que queremos es una “alineación” de los estados mentales, que se individualizan por sus contenidos, con las estructuras neuronales y los eventos que están causalmente involucrados en producir la mentalidad. Podemos llamar al conjunto de tales estados y procesos mentales nuestra “ontología cognitiva” (Anderson). El propósito de la neurociencia cognitiva, entonces, es producir el mapeo que va desde los elementos de nuestra ontología cognitiva a las estructuras neuronales, que a su vez se describen desde las estrategias del diseño y de la física.

El argumento de nuestra sección previa tiene una consecuencia importante aquí. La neurociencia cognitiva, tal como originalmente se concibió, ha sido modelada según los enfoques de descomposición funcional de GOFAI y de arriba hacia abajo. Es decir, los neurocientíficos cognitivos comienzan con modelos de sistemas cognitivos de orden superior, como la memoria, y asumen que estos se descomponen en subsistemas más simples y “más tontos”, como codificadores y recuperadores. Como resultado, muchos neurocientíficos cognitivos diseñan sus experimentos para localizar los mecanismos que actúan

como codificadores durante la codificación, o como recuperadores durante la recuperación. La vieja neurociencia cognitiva, por así decirlo, intenta encontrar los homúnculos tontos en el cerebro.

Pero ahora Dennett sugiere que las estructuras neuronales que buscamos pueden no ser los homúnculos tontos que esperábamos que fueran. Más bien, pueden involucrar un conjunto mucho más complejo de entidades organizadas y conectadas de maneras que no corresponden a los niveles de descomposición de arriba hacia abajo. En consecuencia, las descripciones funcionales de las estructuras neuronales que identifiquemos diferirán dependiendo de si las caracterizamos desde la perspectiva de abajo hacia arriba o desde la de arriba hacia abajo. Pero si este es el caso, ¿cómo debemos interpretar las “activaciones” neuronales medidas en neurociencia? ¿Las interpretamos como realizando cálculos relevantes para una función caracterizada desde la perspectiva de arriba hacia abajo, o para una caracterizada de abajo hacia arriba? ¿Qué hacemos cuando las redes neuronales cuya actividad medimos no corresponden a “codificadores” y “recuperadores”, sino que deberían describirse de formas que den sentido al comportamiento de las neuronas y redes neuronales como agentes?

Una manera más precisa de plantear estas preguntas es la siguiente: ¿qué pasa con la neurociencia cognitiva si la suposición acerca de su objetivo —la existencia de un mapeo limpio (aunque difícil de encontrar) entre nuestra ontología cognitiva y las estructuras neuronales— es defectuosa? ¿Qué pasa si no existe un mapeo limpio, o descubrimos que podemos generar muchos mapeos plausibles pero incompatibles? ¿Deberíamos reconsiderar la neurociencia cognitiva misma, si resulta que la suposición que la impulsa es falsa o mal concebida? ¿Cómo hacemos neurociencia cognitiva en absoluto si hay razones de principio por las cuales no podemos tender puentes entre las descripciones intencionales hechas al nivel de las personas y las hechas al nivel de las neuronas y redes neuronales?

Los cambios en las ideas de Dennett sobre el funcionalismo homuncular plantean preguntas interesantes sobre cómo concibe el mapeo mente-cerebro. Pero también revelan tensiones endémicas a la neurociencia cognitiva misma, pues los problemas acerca de las descripciones intencionales en diferentes niveles de organización están en el centro de nuestros intentos de movernos entre la psicología y la neurociencia. En cierto sentido, las tendencias pragmáticas presentes en otros trabajos de Dennett pueden ser valiosas aquí: quizá todo lo que la neurociencia cognitiva pueda hacer es intentar elaborar el mejor mapeo o conjunto de mapeos posibles, dadas las restricciones filosóficas que hemos discutido. ¿Hace esto de Dennett una especie de instrumentalista o una especie de realista acerca del mapeo mismo?

Conclusión

En su forma temprana, la visión funcionalista homuncular de Daniel Dennett defendía una descomposición funcional de arriba hacia abajo de la cognición. En los niveles superiores de la descomposición, las descripciones de estos estados requerirán lenguaje intencional. Pero al descomponer cada nivel en sus subcomponentes, comenzamos a eliminar las descripciones intencionales, hasta que finalmente llegamos a un nivel “tonto” cuyo comportamiento puede ser reproducido por una máquina siguiendo instrucciones simples.

Una descomposición funcional como esta asume que el nivel inferior realmente carece de mente, y que no necesitamos ningún lenguaje intencional para describir su comportamiento. Es en este punto, sin embargo, donde FBBB sugiere un cambio importante: ya no podemos considerar a las neuronas tan tontas. Es decir, su comportamiento es lo suficientemente complejo como para que explicarlo también requiera adoptar la estrategia intencional. Así, nuestra descomposición funcional de arriba hacia abajo parece alcanzar un nivel en el que el lenguaje intencional reaparece.

Hemos argumentado que esta versión de la visión funcionalista homuncular enfrenta un problema: las descripciones funcionales usadas para captar el comportamiento de las neuronas pueden no encajar bien en la jerarquía anidada de descomposiciones funcionales obtenidas desde un análisis tradicional de arriba hacia abajo. Si este es el caso, entonces mapear los estados y procesos mentales a las estructuras neuronales puede ser difícil o incluso imposible.

FBBB retoma varias ideas que Dennett ha discutido en trabajos previos. Aquí identificamos una de ellas, el funcionalismo homuncular, y mostramos cómo el cambio de opinión de Dennett abre poderosas posibilidades explicativas para las estructuras neuronales. Pero este poder puede tener un precio, pues se vuelve poco claro cómo vincular los diferentes enfoques que podemos adoptar para caracterizar la mente, el cerebro y sus funciones.

Bibliografía

- Anderson, Michael. *After Phrenology: Neural Reuse and the Interactive Brain*. MIT PRESS, 2014.
- Armstrong, David Malet. *A Materialist Theory of Mind*. Routledge, 1968.
- Block, Ned. “Troubles with Functionalism.” *Minnesota Studies in the Philosophy of Science*, vol. 9, 1978, pp. 261–325.
- Caruso, Valeria, et al. “Single Neurons May Encode Simultaneous Stimuli by Switching between Activity Patterns.” *Nature Communications*, vol. 9, 2018, pp. 1–16.

- Clark, Andy. *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. MIT Press, 1989.
- Cummins, Robert. “‘How Does It Work?’ versus ‘What Are the Laws?’: Two Conceptions of Psychological Explanation.” *Explanation and Cognition*. Editado por Frank Keil y Robert Wilson. MIT Press, 2000.
- De Brigard, Felipe. “Review of *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*, Daniel C. Dennett.” *Ideas y Valores*, vol. 56, no. 134, 2007, pp. 129–36.
- . “Predictive Memory and the Surprising Gap: Commentary on Andy Clark’s ‘Whatever Next? Predictive Brains, Situated Agents and the Future of Cognitive Science’.” *Frontiers in Psychology*, vol. 3, no. 420, 2012.
- . *Brainstorms: Philosophical Essays on Mind and Psychology*. Bradford Books, 1978.
- . *Content and Consciousness*. Routledge, 1969.
- Dennett, Daniel. *From Bacteria to Bach and Back: The Evolution of Minds*. W. W. Norton & Company, 2017.
- . *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. MIT Press, 2005.
- Gessell, Bryce. “Indeterminism in the Brain.” *Biology and Philosophy*, vol. 32, no. 6, 2017, pp. 1205–23. [<https://doi.org/10.1007/s10539-017-9601-3>]
- Godfrey-Smith, Peter. “Mind, Matter, and Metabolism.” *The Journal of Philosophy*, vol. 113, 2016, pp. 481–506.
- Lewis, David. “How to Define Theoretical Terms.” *The Journal of Philosophy*, vol. 67, no. 13, 1970, pp. 427–44.
- . “Psychophysical and Theoretical Identification.” *Australasian Journal of Philosophy*, vol. 50, no. 3, 1972, pp. 249–58.
- Lycan, William. *Consciousness*. MIT Press, 1987.
- . “Form, Function, and Feel.” *The Journal of Philosophy*, vol. 78, no. 1, 1981, pp. 24–50.
- Marr, David. “Early Processing of Visual Information.” *Philosophical Transactions of the Royal Society B*, vol. 275, no. 942, 1976, pp. 483–519. [<https://doi.org/10.1098/rstb.1976.0090>]
- McCulloch, Warren, and Walter Pitts. “A Logical Calculus of the Ideas Immanent in Nervous Activity.” *Bulletin of Mathematical Biophysics*, vol. 5, 1943, pp. 115–33. [<https://doi.org/10.1007/BF02478259>]
- Pastuzyn, Elissa, et al. “The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer.” *Cell*, vol. 172, no. 1, 2018, pp. 275–88. [<https://doi.org/10.1016/j.cell.2017.12.024>]
- Putnam, Hilary. “Minds and Machines.” *Dimensions of Mind*. Editado por Sidney Hook. New York University Press, 1960.
- Putnam, Hilary. “Psychological Predicates.” *Art, Mind, and Religion*. Editado por William Capitan y Daniel Merrill. University of Pittsburgh Press, 1967.
- Shallice, Tim. *From Neuropsychology to Mental Structure*. Cambridge University Press, 1988.

Shimizu, Kazumichi, and Mark Stopfer. "Gap Junctions." *Current Biology*, vol. 23, no. 23, 2013. [<https://doi.org/10.1016/j.cub.2013.10.067>]

Simon, Herbert Alexander. *The Sciences of the Artificial*. MIT Press, 1969.

Turing, Alan. "Computing Machinery and Intelligence." *Mind*, vol. 49, 1950, pp. 433–60.