

Andy Clark y David Chalmers,¹ La mente extendida.²

Traducción de **Sergio Quintero Martín**³
Universidad Complutense de Madrid, España
Recibido 31 agosto 2019 · Aceptado 21 febrero 2021

1 · Introducción

¿Dónde se acaba la mente y comienza el resto del mundo? La pregunta invita a dos respuestas estándar. Algunos aceptan los límites de la piel y el cráneo, y afirman que lo que está fuera del cuerpo está también fuera de la mente. Otros, sin embargo, se inclinan por los argumentos que sugieren que el significado de nuestras palabras “simplemente no están en la cabeza”, y sostienen que este externalismo sobre el significado lleva a un externalismo de la mente. Nos proponemos perseguir una tercera posición. Abogamos por

1 N. del T.: Este artículo, publicado en 1998 en la revista *Analysis* (58: 10–23), es uno de los textos más interesantes del debate filosófico en torno a las herramientas tecnológicas y su influjo en la mente humana. Chalmers, filósofo de la Universidad de Arizona, y Clark, a la sazón investigador en Filosofía de la Universidad de Washington en su Departamento de Filosofía, reanudaron con mucha chispa el debate filosófico que, en torno a la cognición, el lenguaje y el entorno social, ocupó a grandes pensadores del siglo XX, como Jean Piaget, Noam Chomsky, Donald Davidson, Hilary Putnam o Tyler Burge. La innovación de Chalmers y Clark fue fundamentar su argumentación sobre la sinergia entre el ser humano y las herramientas digitales innovadoras como el ordenador. Pretenden mostrar cómo las tesis de Putnam en torno a la extensión mental producida por la herramienta del lenguaje pueden verse escenificadas por la extensión mental que supone la inteligencia artificial y los ordenadores de la nueva era. El estilo y la calidad del artículo además abrieron paso a la contemplación completamente moderna de viejos problemas psicológicos, que estaban un tanto en decaída en los últimos tiempos

2 Los autores se enumeran en orden a su grado de creencia en la tesis central.

3 squint46664@gmail.com

un externalismo muy diferente: un *externalismo activo*, basado en el papel activo del entorno en la consecución de los procesos cognitivos.

2 · La cognición extendida

Consideremos tres casos de resolución de problemas humanos:

- (1) Una persona se sienta frente a una pantalla de ordenador que muestra imágenes de varias formas geométricas bidimensionales y se le pide que responda preguntas en relación con el ajuste potencial de esas formas con el “hueco” representado. Para evaluar el ajuste, la persona debe rotar mentalmente las formas para alinearlas con los huecos.
- (2) Una persona se sienta frente a una pantalla similar, pero esta vez puede elegir entre rotar físicamente la imagen en la pantalla, presionando un botón de rotación, o rotar mentalmente las imágenes como antes. También podemos suponer, con bastante verosimilitud, que se produce alguna ventaja de velocidad en la operación de rotación física.
- (3) En algún momento de un futuro *cyberpunk*, una persona está sentada frente a una pantalla de ordenador similar. Este agente, sin embargo, posee el beneficio de un implante neuronal con el que puede realizar la operación de rotación tan rápido como el ordenador en el ejemplo anterior. El agente todavía debe elegir qué recurso interno utilizar (el implante o la eficaz, aunque anticuada, rotación mental), puesto que cada recurso demanda diferente atención y otras actividades cerebrales simultáneas.

¿Cuán presente está la *cognición* en estos casos? Sugerimos que los tres casos son similares. El caso (3) con el implante neuronal parece alinearse claramente con el caso (1). Y el caso (2) con el botón de rotación muestra el mismo tipo de estructura computacional que el caso (3), distribuida entre agente y ordenador en lugar de internalizada dentro del agente. Si la rotación en el caso (3) es cognitiva, ¿con qué derecho consideramos el caso (2) como fundamentalmente diferente? No podemos señalar simplemente el límite piel/cráneo como justificación, ya que la legitimidad de ese límite es precisamente lo que está en discusión. Pero en nada más parecen diferentes.

El tipo de caso que acabamos de describir no es en absoluto tan exótico como pudiera parecer. No se trata solo de la presencia de recursos informáticos externos avanzados lo que plantea el problema, sino de la tendencia general de los razonadores humanos a apoyarse en gran medida en los apoyos ambientales. Por lo tanto, considerar el uso de lápiz y papel para realizar una larga multiplicación (McClelland et al., 1986; Clark, 1989), el uso de reconfiguraciones de piezas físicas con letras para activar la memoria léxica en el juego *Scrabble* (Kirsh, 1995), el uso de instrumentos como la brújula náutica (Hutchins, 1995), y la parafernalia general del lenguaje, los libros, los diagramas, y la cultura. En todos estos casos, el cerebro individual realiza algunas operaciones, algunas de las cuales se delegan a la manipulación de medios externos. Si nuestros cerebros fueran diferentes, esta distribución de las tareas sin duda habría variado.

De hecho, incluso los casos de rotación mental descritos en los escenarios (1) y (2) son reales. Los casos reflejan opciones disponibles para los jugadores del juego de ordenador *Tetris*. En el *Tetris*, las formas geométricas que descienden deben dirigirse rápidamente hacia una ranura apropiada en una estructura emergente. Un botón de rotación puede ser utilizado. David Kirsh y Paul Maglio (1994) calculan que la rotación física de una pieza en 90 grados toma alrededor de 100 milisegundos, más unos 200 milisegundos para apretar el botón. Para lograr el mismo resultado mediante la rotación mental serían necesarios 1000 milisegundos. Kirsh y Maglio continúan presentando una evidencia convincente de que la rotación física se usa no solo para colocar una pieza lista para encajar en una ranura, sino también para ayudar a *determinar* si la forma de la pieza y la ranura son compatibles. Este último uso constituye un caso de lo que Kirsh y Maglio llaman una “acción epistémica”. Las acciones *epistémicas* alteran el mundo con el fin de ayudar y aumentar los procesos cognitivos como el reconocimiento y la búsqueda. Las acciones meramente *pragmáticas*, por el contrario, alteran el mundo porque algún cambio físico es deseable por su propio bien (por ejemplo, poner cemento en un agujero en una presa).

Sugerimos que la acción epistémica exige la propagación del *crédito epistémico*. Si, al abordar alguna tarea, una parte del mundo funciona como un proceso que, *si se hiciera en la cabeza*, no dudaríamos en reconocerlo como parte del proceso cognitivo, entonces esa parte del mundo es (reivindicamos)

parte del proceso cognitivo. ¡Los procesos cognitivos no están (todos) en la cabeza!

3 · El externalismo activo

En estos casos, el organismo humano está vinculado con una entidad externa en una interacción de dos direcciones, creando un *sistema ensamblado* que puede verse como un sistema cognitivo por derecho propio. Todos los componentes en el sistema desempeñan un papel causal activo, y gobiernan conjuntamente el comportamiento de la misma forma en que comúnmente lo hace la cognición. Si eliminamos el componente externo, la competencia conductual del sistema disminuirá, tal como lo haría si elimináramos parte de su cerebro. Nuestra tesis es que este tipo de proceso ensamblado cuenta igualmente como un proceso cognitivo, esté o no completamente en la cabeza.

Este externalismo difiere de la variedad estándar defendida por Putnam (1975) y Burge (1979). Cuando creo que el agua está húmeda y mi gemelo cree que el agua gemela está húmeda, las características externas responsables de la diferencia en nuestras creencias son distales e históricas, situadas en el otro extremo de una larga cadena causal. Los rasgos del *presente* no son relevantes: si estoy rodeado por XYZ en este momento (tal vez me haya teletransportado a la Tierra Gemela), mis creencias aún se refieren al agua estándar, debido a mi historia. En estos casos, las características externas relevantes son *pasivas*. Debido a su naturaleza distal, no desempeñan ningún papel en la dirección del proceso cognitivo en el aquí y el ahora. Esto se refleja en el hecho de que las acciones realizadas por mí y mi gemelo son físicamente indistinguibles, a pesar de nuestras diferencias externas.

En los casos que describimos, en contraste, las características externas relevantes están *activas*, desempeñando un papel crucial en el aquí y el ahora. Debido a que están ensamblados con el organismo humano, tienen un impacto directo en el organismo y en su comportamiento. En estos casos, las partes relevantes del mundo están *en el ajo*, no colgando en el otro extremo de una larga cadena causal. Concentrarse en este tipo de ensamblaje nos lleva a un *externalismo activo*, en oposición al externalismo pasivo de Putnam y Burge.

Muchos objetan que incluso si Putnam y Burge tienen razón sobre la externalidad del contenido, no está claro que estos aspectos externos desempeñen un papel causal o explicativo en la generación de la acción. En los casos contrafácticos en los que la estructura interna se mantiene constante, pero estas características externas se modifican, el comportamiento se mantiene igual; así, la estructura interna parece estar haciendo el trabajo crucial. Aquí no resolveremos este problema, pero observamos que el externalismo activo no se ve amenazado por ningún problema de este tipo. Las características externas en un sistema ensamblado desempeñan un papel ineliminable –si mantenemos la estructura interna, pero cambiamos las características externas, el comportamiento puede cambiar completamente. Las características externas aquí son tan causalmente relevantes como las características internas típicas del cerebro⁴.

Al abrazar el externalismo activo, permitimos una explicación más natural de todo tipo de acciones. Entonces puedo explicar mi elección de palabras en el *Scrabble*, por ejemplo, como el resultado de un proceso cognitivo extendido que involucra la reconfiguración de las piezas en mi bandeja. Por supuesto, siempre se podría tratar de explicar mi acción en términos de procesos internos y junto con una serie de “inputs” y “acciones”, pero esta explicación sería innecesariamente compleja. Si tuviera un proceso isomórfico en la cabeza, no sentiríamos ninguna necesidad de caracterizarlos de esta manera. En un sentido muy real, la reconfiguración de las piezas en mi bandeja no es parte de la acción; es parte del *pensamiento*.

La visión que defendemos aquí se refleja en un creciente cuerpo de investigaciones en ciencia cognitiva. En áreas tan diversas como la teoría de la cognición situada (Suchman, 1987), los estudios de robótica y mundo real (Beer, 1989), los enfoques dinámicos para el desarrollo infantil (Thelen

⁴ Gran parte del atractivo del externalismo en la filosofía de la mente puede provenir del atractivo intuitivo del externalismo activo. Los externalistas a menudo hacen analogías con características externas en sistemas ensamblados, y apelan a la arbitrariedad de los límites entre el cerebro y el entorno. Pero estas intuiciones no terminan de encajar en la inscripción del externalismo estándar. En la mayoría de los casos de Putnam/Burge, el entorno inmediato es irrelevante; solamente el entorno histórico cuenta. El debate se ha centrado en la cuestión de si la mente debe estar en la cabeza, pero una pregunta más relevante al evaluar estos ejemplos podría ser: ¿está la mente en el presente?

y Smith, 1994), y la investigación sobre las propiedades cognitivas de los colectivos (Hutchins, 1995), la cognición se considera a menudo que es continua con los procesos en el entorno⁵. Por lo tanto, ver la cognición como un proceso extendido no es simplemente una decisión terminológica; establece una diferencia significativa en la metodología de la investigación científica. En efecto, los métodos explicativos que una vez podrían haber sido considerados apropiados solo para el análisis de los procesos “internos” ahora se están adaptando para el estudio de lo externo, y existe la esperanza de que nuestra comprensión de la cognición se enriquecerá.

Algunos encuentran este tipo de externalismo inaceptable. Una razón puede ser que muchos identifican lo cognitivo con lo consciente, y parece lejos de ser plausible que la conciencia se extienda fuera de la cabeza en estos casos. Pero no todos los procesos cognitivos, al menos en el uso estándar, son procesos conscientes. Se acepta ampliamente que todo tipo de procesos más allá de los límites de la conciencia juegan un papel crucial en el procesamiento cognitivo: en la recuperación de recuerdos, en los procesos lingüísticos, y en la adquisición de habilidades, por ejemplo. Entonces, el mero hecho de que los procesos externos son externos cuando la conciencia es interna no es razón para negar que esos procesos sean cognitivos.

Más interesante, uno podría argumentar que lo que mantiene en la cabeza los procesos de cognición reales es el requisito de que los procesos de cognición sean *portátiles*. Aquí, nos mueve una visión de lo que podríamos llamar la Mente Desnuda: un conjunto de recursos y operaciones que siempre podemos aportar en una tarea cognitiva, sin importar del entorno local. Desde este punto de vista, el problema con los sistemas ensamblados es que se *desensamblan* fácilmente. Los verdaderos procesos cognitivos serían aquellos que se encuentran en el núcleo constante del sistema; cualquier otra cosa es un complemento adicional.

Existe algo en esta objeción. El cerebro (o cerebro y cuerpo) comprende un conjunto de recursos cognitivos básicos, portátiles y de interés por sí mismos. Estos recursos pueden incorporar acciones corporales en los procesos cognitivos, como cuando usamos nuestros dedos como memoria de trabajo en un cálculo complicado, pero no abarcarán los aspectos más

⁵ Se pueden encontrar puntos de vista filosóficos de un espíritu similar en Haugeland (1995), McClamrock (1995), Varela et al. (1991) y Wilson (1994).

contingentes de nuestro entorno externo, como una calculadora de bolsillo. Aun así, la mera contingencia del ensamblaje no descarta el estado cognitivo. En un futuro lejano, podremos conectar varios módulos en nuestro cerebro para ayudarnos: un módulo para memoria adicional a corto plazo cuando lo necesitemos, por ejemplo. Cuando se conecte el módulo, los procesos que lo involucran son tan cognitivos como si hubieran estado allí todo el tiempo.

Incluso si uno hiciera que el criterio de portabilidad fuera fundamental, el externalismo activo no se vería afectado. Contar con nuestros dedos siempre se ha considerado algo aceptado en el proceso cognitivo, por ejemplo, y es fácil llevar las cosas más allá. Pensemos en la vieja imagen del ingeniero con una cinta métrica que cuelga de su cinturón donde quiera que va. ¿Qué pasaría si las personas siempre llevaran una calculadora de bolsillo, o se la hubieran implantado? El criterio verdadero de la intuición de la portabilidad es que para que los sistemas ensamblados sean relevantes para el núcleo de la cognición, se requiere un ensamblaje *fiabile*. Sucede que el ensamblaje más fiable tiene lugar dentro del cerebro, pero también puede aparecer un ensamblaje fiable ligado al entorno. Si los recursos de mi calculadora y mi agenda *Filofax* están siempre disponibles cuando los necesito, entonces están ensamblados conmigo de la manera más fiable que necesitamos. En efecto, son parte del conjunto básico de recursos cognitivos que llevo encima en el mundo cotidiano. Estos sistemas no pueden impugnarse simplemente por el peligro de un daño concreto, pérdida o mal funcionamiento, o por algún desensamblaje ocasional: el cerebro biológico corre un peligro similar y, en ocasiones, pierde temporalmente las capacidades en episodios de sueño, de intoxicación y de emoción. Si las capacidades relevantes están generalmente donde son requeridas, entonces esto es un ensamblado suficiente.

Además, puede ser que el cerebro biológico haya evolucionado y madurado de manera tal que tenga en cuenta la presencia fiable de un entorno externo manipulable. Ciertamente, parece que la evolución ha favorecido las capacidades de almacenamiento que han sido especialmente orientadas a parasitar el entorno local para reducir la carga de memoria, e incluso para transformar la naturaleza de los problemas computacionales en sí mismos. Nuestros sistemas visuales han evolucionado para depender de su entorno de varias maneras: aprovechan los hechos contingentes derivados de la estructura de escenas naturales (Cf. Ullman y Richards, 1994), por ejemplo, y se

benefician de los atajos computacionales proporcionados por el movimiento corporal y locomoción (Cf. Blake y Yuille, 1992). Quizás haya otros casos en los que la evolución ha encontrado ventajoso explotar la posibilidad de que el entorno esté integrado en la cognición. Si es así, entonces el ensamblaje externo es parte del conjunto verdaderamente básico de recursos cognitivos que aportamos al mundo.

Otro ejemplo puede ser el lenguaje, que parece ser un medio central por el cual los procesos cognitivos se extienden al mundo. Pensemos en un grupo de personas reunidas en una mesa para generar ideas, en un filósofo que piensa mejor escribiendo, desarrollando sus ideas a medida que avanza. Puede ser que el lenguaje haya evolucionado, en parte, para habilitar tales extensiones de nuestros recursos cognitivos dentro de sistemas ensamblados activamente.

También dentro del tiempo de vida de un organismo, el aprendizaje individual puede haber moldeado el cerebro de maneras que dependen de las extensiones cognitivas que le han rodeado conforme iba aprendiendo. El lenguaje es nuevamente un ejemplo central aquí, al igual que los diversos artefactos físicos y computacionales que se utilizan habitualmente como extensiones cognitivas por parte de los niños en las escuelas y por quienes se forman en numerosas profesiones. En tales casos, el cerebro se desarrolla de una manera que complementa las estructuras externas y aprende a desempeñar su papel dentro de un sistema unificado y densamente ensamblado. Una vez que reconocemos que el papel crucial del entorno para forzar la evolución y el desarrollo cognitivo, vemos que la cognición extendida es un proceso cognitivo central, no un complemento adicional.

4 • De la cognición a la mente

Hasta ahora hemos hablado en gran parte sobre el “proceso cognitivo”, y hemos argumentado sobre su extensión en el entorno. Algunos podrían pensar que la conclusión se ha alcanzado con muy poco esfuerzo. Tal vez algún *procesamiento* tenga lugar en el entorno, pero ¿de qué *mente*? Todo cuanto hemos dicho hasta ahora es compatible con la perspectiva de que los estados verdaderamente cerebrales (las experiencias, las creencias, los deseos, las

emociones, etc.) están determinados por los estados del cerebro. Tal vez, lo que es verdaderamente mental sea interno, ¿después de todo?

Nos proponemos llevar las cosas un paso más allá. Mientras que algunos estados mentales, como las experiencias, tal vez determinados internamente, hay otros casos en los que factores externos hacen una contribución significativa. En particular, argumentaremos que las *creencias* pueden estar constituidas en parte por características del entorno, cuando esas características desempeñan el papel apropiado en la conducción de los procesos cognitivos. Si es así, la mente se extiende al mundo.

Primero, consideremos un caso de creencia implantada en la memoria. Inga oye decir a un amigo que hay una exposición en el Museo de Arte Moderno y decide ir a verla. Ella piensa por un momento y recuerda que el museo está en la calle 53 porque lo que camina a hasta esa calle y entra en el museo. Parece claro que Inga cree que el museo está en la calle 53 y lo creía incluso antes de consultar su memoria. Anteriormente, no era una creencia *sobreenvenida*, pero tampoco la mayoría de nuestras creencias. La creencia estaba en algún lugar en la memoria, esperando para que se accediera a ella.

Ahora consideremos a Otto. Otto padece la enfermedad de Alzheimer y, como muchos pacientes de Alzheimer, confía en la información del entorno para ayudar a estructurar su vida. Otto lleva a todas partes consigo un cuaderno. Cuando aprende nueva información, la escribe. Cuando necesita alguna información antigua, la busca. Para Otto, su cuaderno desempeña el papel que generalmente desempeña una memoria biológica. Hoy, Otto se entera de la exposición en el Museo de Arte Moderno y decide ir a verla. Consulta el cuaderno, que dice que el museo está en la calle 53, así que camina hasta esa calle y entra en el museo.

Claramente, Otto caminó a la calle 53 porque quería ir al museo y creía que el museo estaba en la calle 53. Y justo como Inga tenía su creencia incluso antes de consultar su memoria, parece razonable decir que Otto creía que el museo estaba en la calle 53 incluso antes de consultar su cuaderno. En los aspectos relevantes, los casos son completamente análogos: el cuaderno juega para Otto el mismo papel que juega la memoria para Inga. La información en el cuaderno funciona igual que la información que constituye una creencia ordinaria no *sobreenvenida*; simplemente sucede que esta información se encuentra más allá de la piel.

La alternativa es decir que Otto no tiene ninguna creencia sobre el asunto hasta que consulta su cuaderno; en el mejor de los casos, él cree que el museo está ubicado en la dirección indicada en el cuaderno. Pero si seguimos a Otto por un tiempo, veremos cuán antinatural es esta forma de hablar. Otto está empleando constantemente el cuaderno en su rutina. Es fundamental para sus acciones en todo tipo de contextos, en la forma en que una memoria ordinaria es central en una vida cotidiana. La misma información podría aparecer una y otra vez, tal vez modificada ligeramente en alguna ocasión, antes de retirarse a los rincones de su memoria artificial. Decir que las creencias desaparecen cuando se archiva el cuaderno parece pasar por alto el panorama general, al igual que decir que las creencias de Inga desaparecen tan pronto como ella se da cuenta de ellas. En ambos casos, la información es fiable allí donde se la necesita, está disponible para la conciencia y está disponible para guiar la acción, de manera que esperamos que sea una creencia.

Ciertamente, en la medida en que las creencias y los deseos se caracterizan por sus roles explicativos, los casos de Otto e Inga parecen estar a la par: la dinámica causal esencial de los dos casos se refleja entre sí con precisión. Nos satisface explicar la acción de Inga en términos de su deseo sobrevenido de ir al museo y su creencia permanente de que el museo está en la calle 53, y nos convence explicar la acción de Otto de la misma manera. La alternativa es explicar la acción de Otto en términos de su deseo de ir al museo, su creencia permanente de que el Museo está en la ubicación escrita en el cuaderno, y el hecho accesible de que el cuaderno dice que el Museo está en la calle 53; pero eso complica la explicación innecesariamente. Si debemos recurrir a explicar la acción de Otto de esta manera, entonces debemos hacerlo por las incontables acciones en las que está involucrado su cuaderno; en cada una de las explicaciones, habrá un término adicional relacionado con el cuaderno. Creemos que explicar las cosas de este modo es *llevarlas demasiado lejos*. Es inútilmente complejo, de la misma manera que sería inútilmente complejo explicar las acciones de Inga en términos de creencias sobre su memoria. El cuaderno es una constante para Otto, de la misma manera que la memoria es una constante para Inga; señalarlo en cada creencia/deseo sería redundante. Es una explicación, la simplicidad es poder.

Si esto es correcto, podemos incluso construir el caso de un Otto gemelo, que es como Otto, excepto que hace un tiempo escribió erróneamente

en su cuaderno que el Museo de Arte Moderno estaba en la calle 51. Hoy, el Otto gemelo es un duplicado físico de Otto de piel para adentro, pero su cuaderno difiere. En consecuencia, el Otto gemelo se caracteriza mejor por creer que el museo está en la calle 51, donde Otto cree que está en la calle 53. Es estos casos, una creencia simplemente no está en la cabeza.

Esto refleja la conclusión de Putnam y Burge, pero nuevamente hay diferencias importantes. En los casos de Putnam/Burge, las características externas que constituyen diferencias en la creencia son distales e históricas, de modo que los gemelos en estos casos producen físicamente una conducta distinguible. En los casos que describimos, las características externas relevantes desempeñan un papel activo en el aquí y el ahora, y tienen un impacto directo en el comportamiento. Donde Otto camina hasta la calle 53, el Otto gemelo va hasta la calle 51. No hay duda de la irrelevancia explicativa para este tipo de contenido de creencias externas; se introduce precisamente por el papel explicativo central que desempeña. Al igual que los casos de Putnam y Burge, estos casos implican diferencias en la referencia y en las condiciones de verdad, pero también implican diferencias en la dinámica de la *cognición*⁶.

La conclusión es que cuando se trata de creer, no hay nada sagrado en el cráneo y la piel. Lo que hace que cierta información cuente como una creencia es el rol que desempeña, y no hay ninguna razón por la que el rol relevante solo se puede desempeñar desde dentro del cuerpo.

Algunos se resistirán a esta conclusión. Un oponente puede salir a escena e insistir en que, al usar el término “creencia”, o tal vez incluso de acuerdo con el uso estándar, Otto simplemente no puede estar creyendo que el museo está en la calle 53. No pretendemos debatir qué es el uso estándar; nuestra perspectiva amplía la noción de creencia que *debería* usarse cuando Otto está teniendo la creencia en cuestión. En todos los aspectos *importantes*, el caso de Otto es similar a un caso estándar de creencia (no sobrevenida). Las diferencias entre el caso de Otto y el de Inga son sorprendentes, pero

⁶ En la terminología de Chalmers: los gemelos en los casos de Putnam y Burge difieren solo en su contenido relacional (intensión secundaria), pero se puede ver que Otto y su gemelo difieren en su contenido nocional (intensión primaria), que es el tipo de contenido que gobierna la cognición. El contenido nocional generalmente es interno en un sistema cognitivo, pero en este caso el sistema cognitivo se extiende efectivamente para incluir el cuaderno [N. del T.: Cf. Chalmers, 2002].

son superficiales. Al utilizar la noción de “creencia” de una manera más amplia, resulta algo más parecido a una forma natural. La noción se vuelve más profunda y más unificada, y es más útil en la explicación.

Para proporcionar una resistencia sustancial, un oponente tiene que demostrar que los casos de Otto e Inga difieren en algunos aspectos importantes y relevantes. Pero ¿en qué aspectos profundos son diferentes los casos? Argumentar únicamente sobre la base de que la información está en la cabeza es un caso, pero no en el otro no sería sólido. Si esta diferencia es relevante para una diferencia en la creencia, seguramente no es *primitivamente* relevante. Para justificar el tratamiento diferente, debemos encontrar una diferencia más básica subyaciendo entre los dos.

Se podría sugerir que los casos son significativamente diferentes, ya que Inga tiene un acceso más *fiable* a la información. Después de todo, alguien podría quitarle el cuaderno a Otto en cualquier momento, por lo que la memoria de Inga es más segura. No es inverosímil que la constancia sea relevante: en efecto, el hecho de que Otto siempre use su cuaderno desempeñó un papel en nuestra justificación de su estado cognitivo. Si Otto consultara una guía de manera única, sería mucho menos probable que le atribuyamos una creencia permanente. Pero en el caso original, el acceso de Otto al cuaderno es muy fiable —no es perfectamente fiable, por cierto, pero tampoco lo es el acceso de Inga a su memoria. Una cirugía podría alterar su cerebro, o más coloquialmente, podría ser que hubiera bebido mucho. La mera posibilidad de tal manipulación no es suficiente para negarle la creencia.

Uno podría preocuparse de que el acceso de Otto a su cuaderno *de hecho* va y viene. Se ducha sin el cuaderno, por ejemplo, y no puede leerlo cuando está oscuro. ¿Seguramente su creencia no puede ir y venir tan fácilmente? Podríamos solucionar este problema redescribiendo la situación, pero, en cualquier caso, una desconexión temporal ocasional no amenaza nuestra afirmación. Después de todo, cuando Inga está dormida, o cuando está intoxicada, no decimos que su creencia desaparece. Lo que realmente cuenta es que la información está fácilmente disponible cuando el sujeto la necesita, y esta restricción se satisface por igual en los dos casos. Si el cuaderno de Otto a menudo no estuviera disponible para él en momento en que la información que contiene fuera útil, podría haber un problema, ya que la información no podría desempeñar el papel de guía de la acción que

es fundamental para la creencia; pero si está fácilmente disponible en las situaciones más relevantes, la creencia no está en peligro.

Quizá una diferencia es que Inga tiene *mejor* acceso a la información que Otto. Los procesos “centrales” de Inga y su memoria probablemente tiene un enlace de ancho de banda relativamente alto entre ellos, en comparación con la conexión de bajo grado entre Otto y su cuaderno. Pero esto solo no hace una diferencia entre creer y no creer. Consideremos a Lucy, la amiga del museo de Inga, cuya memoria biológica solo tiene un enlace de baja calidad con sus sistemas centrales, debido a una biología no estándar o desventuras pasadas. El procesamiento en el caso de Lucy puede ser menos eficiente, pero siempre que la información relevante sea accesible, Lucy cree claramente que el museo está en la calle 53. Si la conexión fuera demasiado indirecta — si Lucy tuviera que esforzarse para recuperar la información con resultados confusos, o si necesitara la ayuda de un psicoterapeuta— podríamos ser más reacios a atribuir la creencia, pero tales casos están mucho más allá de la situación de Otto, en la que la información es fácilmente accesible.

Otra sugerencia podría ser que Otto tiene acceso a la información relevante solo por la *percepción*, mientras que Inga tiene un acceso más directo, tal vez por medio de la introspección. De alguna manera, sin embargo, poner las cosas de esta manera no sería sólido. Después de todo, en efecto abogamos por una perspectiva sobre el cual los procesos internos de Otto y su cuaderno constituyen un sistema cognitivo único. Desde el punto de vista de este sistema, que el flujo de información entre el cuaderno y el cerebro no es perceptivo en absoluto; no implica el impacto de algo fuera del sistema. Es más parecido al flujo de información dentro del cerebro. La única forma profunda en que el acceso es perceptivo es que, en el caso de Otto, existiera una fenomenología perceptiva asociada con la recuperación de la información, mientras que en el caso de Inga no la haya. Pero ¿por qué la naturaleza de una fenomenología asociada debe hacer una diferencia en el estado de una creencia? La memoria de Inga puede tener alguna fenomenología asociada, pero todavía es una creencia. La fenomenología no es visual, por cierto. Pero para una fenomenología visual, consideremos a *Terminator*, de la película homónima de Arnold Schwarzenegger. Cuando recuerda alguna información de la memoria, se “muestra” ante él en su campo visual (presumiblemente, es consciente de ello, ya que hay tomas frecuentes que representan su

punto de vista). El hecho de que los recuerdos permanentes se recuerden de este modo tan inusual no crea poca diferencia en su condición de creencias permanentes.

Estas pequeñas diferencias entre los casos de Otto e Inga son todas diferencias *superficiales*. Centrarse en ellas sería perder el modo en que, para Otto, las entradas del cuaderno juegan exactamente el mismo papel que las creencias desempeñan en la guía la vida de la mayoría de las personas.

Quizá la intuición de que para Otto no es una creencia verdadera proviene de un sentimiento residual de que las únicas creencias verdaderas son las creencias que ocurren. Si tomamos este sentimiento en serio, la creencia de Inga también se descartará, al igual que muchas creencias que atribuimos en la vida cotidiana. Esta sería una visión extrema, pero puede ser la forma más consistente de negar la creencia de Otto. Incluso desde un punto de vista un poco menos extremo, la opinión de que una creencia debe estar disponible para la conciencia, por ejemplo, la entrada del cuaderno de Otto parece calificar tan bien como la memoria de Inga. Una vez que se dejan entrar las creencias disposicionales, es difícil resistirse a la conclusión de que el cuaderno de Otto tiene todas las disposiciones relevantes.

5 • Más allá de los límites exteriores

Si esta tesis es aceptada, ¿hasta dónde debemos llevarla? Todo tipo de casos paradójicos se nos vienen a la mente. ¿Qué hay de los aldeanos amnésicos en *Cien años de soledad*, que olvidan los nombres de todo y cuelgan etiquetas en todas partes? ¿La información en mi agenda *Filofax* cuenta como parte de mi memoria? Si el cuaderno de Otto ha sido manipulado, ¿creerá en la información recién instalada? ¿Creo en los contenidos de la página que tengo delante de mí antes de leerla? ¿Mi estado cognitivo se extiende de alguna manera a través de Internet?

No creemos que haya respuestas categóricas a todas estas preguntas, y no las daremos. Pero para ayudar a entender lo que está involucrado en las atribuciones de creencias extendidas, podemos al menos examinar las características de nuestro caso central que hacen que la noción sea tan claramente aplicable. Primero, el cuaderno es una constante en la vida de Otto —en los casos en que la información del cuaderno sea relevante, rara

vez actuará sin consultarlo. En segundo lugar, la información en el cuaderno está directamente disponible sin dificultad. Tercero, al recuperar información del cuaderno, él automáticamente la asume. Cuarto, la información en el cuaderno ha sido respaldada, conscientemente en algún momento en el pasado, y de hecho está ahí como consecuencia de este respaldo⁷. El estado de la cuarta característica como un criterio para la creencia es discutible (¿quizá uno puede adquirir creencias a través de la percepción subliminal, o a través de la manipulación de la memoria?), pero las primeras tres características ciertamente juegan un papel crucial.

En la medida en que los casos paradójicos cada vez sean más exóticos carecerán de estas características, la aplicabilidad de la noción de “creencia” se reducirá gradualmente. Si rara vez tomo medidas relevantes sin consultar mi agenda *Filofax*, por ejemplo, su estado dentro de mi sistema cognitivo se parecerá al del cuaderno de Otto. Pero si a menudo actúo sin consultar —por ejemplo, si a veces respondo preguntas relevantes con “no sé” — entonces la información en él contará con menos claridad como parte de mi sistema de creencias. Internet es probable que no sea fiable en varios cargos, a menos que extrañamente sea dependiente del ordenador, esté adaptado a la tecnología y sea confiado, pero la información en ciertos archivos de mi ordenador sí lo será. En casos intermedios, la pregunta de si una creencia está presente puede ser indeterminada, o la respuesta puede depender de los diferentes estándares que están en juego en diversos contextos en los que se plantee la pregunta. Pero cualquier indeterminación aquí no significa que, en los casos centrales, la respuesta no sea clara.

¿Qué sucede con la cognición socialmente extendida? ¿Podrían mis estados mentales estar parcialmente contruidos por los estados de otros pensadores? No vemos ninguna razón para que no sea así, en principio. En una pareja inusualmente dependiente, es totalmente posible que las creencias de un compañero jueguen el mismo tipo de rol para el otro como lo hace

⁷ Los criterios de constancia y respaldo pasado pueden sugerir que la historia es en parte constructiva de la creencia. Uno podría reaccionar a esto eliminando cualquier componente histórico (dando una lectura puramente disposicional del criterio de constancia y eliminando el criterio de la incorporación pasada, por ejemplo), o podría permitir dicho componente siempre que la carga principal la tengan las características en el presente.

el cuaderno para Otto. Lo que es central es el alto grado de confianza, dependencia y accesibilidad. En otras relaciones sociales, es posible que estos criterios no se cumplan con tanta claridad, pero que sin embargo puedan cumplirse en campos específicos. Por ejemplo, el camarero de mi restaurante favorito podría actuar como un repositorio de mis creencias sobre mis comidas favoritas (esto podría incluso interpretarse como un caso de deseo extendido). En otros casos, las creencias de uno pueden estar incorporadas en su secretario, su contable o su colaborador.

En cada uno de estos casos, la mayor carga del ensamblaje entre los agentes es el lenguaje. Sin el lenguaje, podríamos ser mucho más afines a las mentes “internas”, cartesianas y discretas, en las que la cognición de alto nivel se basa en gran medida en los recursos internos. Pero el advenimiento del lenguaje nos ha permitido difundir esta carga en el mundo. El lenguaje, así interpretado, no es un espejo de nuestros estados mentales, sino un complemento de ellos. Sirve como una herramienta cuya función es extender la cognición de manera que los dispositivos a bordo no pueden hacerlo. En efecto, puede ser que la explosión intelectual en la reciente época evolutiva se deba en gran medida a esta extensión lingüística hecha posible como cualquier desarrollo independiente en nuestros recursos cognitivos internos.

¿Qué sucede, finalmente, con el yo? ¿Implica la mente extendida un ser extendido? Así parece. La mayoría de nosotros ya aceptamos que el yo supera los límites de la conciencia; mis creencias disposicionales, por ejemplo, constituyen en cierto sentido parte de lo que soy. Si es así, entonces estos límites pueden también ir más allá de la piel. La información en el cuaderno de Otto, por ejemplo, es una parte central de su identidad como agente cognitivo. De lo que se trata es que *Otto mismo* es mejor entendido como un sistema extendido, un ensamblaje de organismo biológico y de recursos externos. Para resistir constantemente esta conclusión, tendríamos que encoger al yo en un mero conjunto de estados presentes, amenazando gravemente su profunda continuidad psicológica. Es mucho mejor seguir una visión más amplia, y ver a los propios agentes como extensiones del mundo.

Al igual que sucede con cualquier reconcepción de nosotros mismos, esta visión tendrá consecuencias significativas. Hay consecuencias obvias para las visiones filosóficas de la mente y para la metodología de la investigación en ciencia cognitiva, pero también habrá efectos en los campos de la

moral y de lo social. Puede ser, por ejemplo, que, en algunos casos, interferir con el entorno de una persona tenga el mismo significado moral que intervenir con su persona. Y si esta perspectiva se toma en serio, ciertas formas de actividad social puedan ser concebidas como menos afines a la comunicación y la acción, y como más próximas al pensamiento. En cualquier caso, una vez que se ha usurpado la hegemonía de la piel y el cráneo, podemos vernos a nosotros mismos como auténticos seres del mundo.

6 · Referencias bibliográficas

- Beer, R.: *Intelligence as Adaptive Behavior*. Nueva York: Academic, 1989.
- Blake, A; y Yuille, A. (eds.): *Active Vision*. Cambridge: MIT, 1992.
- Burge, T.: "Individualism and the mental" en *Midwest Studies in Philosophy*, 4, 1979, pp. 73-122.
- Chalmers, D. J.: *The components of content*. Chalmers, D. J. (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University, 2002.
- Clark, A.: *Microcognition*. Cambridge: MIT, 1989.
- Haugeland, J.: *Mind embodied and embedded*. Houg, Y.; y Ho, J. (eds.), *Mind and Cognition*. Taipei: Academia Sinica, 1995.
- Hutchins, E.: *Cognition in the Wild*. Cambridge: MIT, 1995.
- Kirsh, D.: "The intelligent use of space" en *Artificial Intelligence*, 73, 1995, pp. 31-68.
- Kirsh, D.; y Maglio, P.: "On distinguishing epistemic from pragmatic action" en *Cognitive Science*, 18, 1994, pp. 513-549.
- McClamrock, R.: *Existential Cognition*. Chicago: University of Chicago, 1995.
- McClelland, J. L.; Rumelhart, D. E.; y Hinton, G. E.: *The appeal of parallel distributed processing*. McClelland, J. L.; y Rumelhart, D. E. (eds.), *Parallel Distributed Processing (Vol. 2)*. Cambridge: MIT, 1986.
- Putnam, H.: *The meaning of 'meaning'*. Gunderson, K. (ed.), *Language, Mind, and Knowledge*. Minneapolis: University of Minnesota, 1975.
- Suchman, L.: *Plans and Situated Actions*. Cambridge: Cambridge University, 1987.
- Thelen, E.; y Smith, L.: *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge: MIT, 1994.

- Ullman, S.; y Richards, W.: *Image Understanding*. Norwood: Ablex, 1984.
- Varela, F.; Thompson, E.; y Rosch, E.: *The Embodied Mind*. Cambridge: MIT, 1991.
- Wilson, R.: “Wide computationalism” en *Mind*, 103, 1994, pp. 351-372.