

# *Incorporización y la filosofía de la mente*\*

ANDY CLARK  
*Universidad de Sussex*

## I. INTRODUCCIÓN: EL REDESCUBRIMIENTO DEL CUERPO Y DEL MUNDO.

LA CIENCIA COGNITIVA es en algún sentido la ciencia de la mente. Pero en los últimos años un asunto crecientemente influyente ha sido el del papel del cuerpo físico, y el del entorno local, para afrontar el éxito adaptativo. Desde luego, ningún científico cognitivo en sus cabales ha mantenido nunca que el cuerpo y el mundo fuesen completamente irrelevantes para entender la mente. Pero había no obstante una tendencia inequívoca a marginar tales factores: insistir en la complejidad interior mientras se simplifica o ignora la compleja intersección interior-exterior que caracteriza el grueso de la solución biológica básica de problemas<sup>1</sup>. Esta tendencia se expresó, por ejemplo, en el desarrollo de los algoritmos de planificación que tratan la acción en el mundo real como un mero modo de implementar soluciones a las que se llega por pura cognición (trabajos más recientes, por el contrario, permiten que tales

\* Traducción de Pedro J. Chamizo © Cambridge University Press.

<sup>1</sup> Excepciones notables a esta tendencia son trabajos tales como J. J. Gibson, *The Ecological Approach to Visual Perception* (Boston: Houghton-Mifflin, 1979) y, en clave más filosófica, *La Structure du Comportement*, de Maurice Merleau-Ponty (Paris: Presses Universitaires de France, 1942). El reciente trabajo sobre la Visión Animada y la óptica ecológica (ver Sección II más adelante) está influido claramente por las ideas gibsonianas; mientras que tratamientos tales como el de F. Varela, E. Thompson y E. Rosch, *The Embodied Mind* (Cambridge: MA: MIT Press, 1991) se confiesan explícitamente seguidores de Merleau-Ponty. Hay una breve discusión de estas raíces históricas en el capítulo 8 de mi *Being There: Putting Brain, Body and World Together* (Cambridge, MA: MIT Press, 1997).

acciones jueguen importantes papeles computacionales y para resolver problemas)<sup>2</sup>. Esto aflora también en la descripción de David Marr<sup>3</sup> de la tarea de la visión como la construcción de una detallada imagen tridimensional de la escena visual. El hecho de poseer un tan rico modelo interior permite al sistema «desechar» el mundo y enfocar posteriormente la actividad computacional solamente hacia el modelo interior<sup>4</sup>. Y más generalmente, la entera visión de la cognición como operaciones interiores sobre modelos internos del mundo refleja una estrategia explicativa que podría ser razonablemente calificada de aislacionismo<sup>5</sup>.

### (Aislacionismo)

El mundo es (precisamente) una fuente de entradas y una cancha para las salidas, y el cuerpo es (precisamente) un órgano para recibir las entradas y efectuar las salidas (acciones). La tarea del procesamiento primerizo es convertir las entradas en un modelo interno del mundo de suficiente riqueza como para permitir al grueso de la actividad resolutoria de problemas ser definida sólo sobre el modelo interior.

El aislacionismo, es justo decirlo, está adquiriendo una creciente mala fama. Pero la forma precisa de una tentativa alternativa permanece oscura. Las tesis antiaislacionistas fluctúan desde la insistencia relativamente inocente en que no conseguiremos una visión equilibrada de lo que hace el cerebro hasta que no prestemos más atención a los complejos papeles de cuerpo y mundo, a la acusación autoconscientemente revolucionaria de que la mente misma no es, después de todo, una especial cancha interior poblada de modelos internos y representaciones, sino, más bien, la operación de un sistema profundamente entretejido, que incorpora aspectos del cerebro, cuerpo y mundo —un sistema que resiste

2 Ver, por ejemplo, P. Agre y S. Rosenschein (eds.) *Computational Theories of Interaction and Agency* (Cambridge, MA: MIT Press, 1996); D. Kirsh y P. Maglio "On Distinguishing Epistemic from Pragmatic Action", en *Cognitive Science* 18 (1995), 513-49; y E. Hutchins, *Cognition in the Wild* (Cambridge, MA: MIT Press, 1995).

3 Ver D. Marr *Vision* (San Francisco, CA: W. H. Freeman, 1982).

4 Esta tradición ha sido suficientemente criticada en P. S. Churchland, V. Ramachandran y T. Sejnowski, "A Critique of Pure Vision", en C. Koch y J. Davies (eds.) *Large-Scale Neuronal Theories of the Brain* (Cambridge, MA: MIT Press, 1994).

5 Los especialistas en robótica aluden (de forma despectiva habitualmente) a esta visión aislacionista como la idea de un Ciclo lineal Sentido-Pensamiento-Acto. Ver, por ejemplo, C. Malcolm, T. Smithers y J. Hallam, "An Emerging Paradigm in Robot Architecture", *Edinburgh University Department of Artificial Intelligence Technical Report*, 1989.

el análisis informativo en términos de las viejas nociones de modelos interiores, representaciones y computación<sup>6</sup>. La visión antiaislacionista más radical describe entonces a los seres humanos como una especie de (por así decirlo) *agente postcartesiano*<sup>7</sup>. El agente postcartesiano es un lugar de conocimiento que obra por razones y tiene creencias y deseos. Sin embargo no aloja representaciones internas y resiste el análisis en términos de cualesquiera distinciones cognitivamente importantes entre los procesos internos y externos, entre la percepción, la cognición y la acción o entre la mente, el cuerpo y el mundo.

Mantendré que la visión postcartesiana no es convincente y que una jugada maestra del argumento (una jugada que apodo «movimiento de lo cognitivo a lo ejecutivo») es dialécticamente sospechosa y empíricamente defectuosa. Dicho en positivo, mantendré una tesis más débil aunque todavía antiaislacionista; una tesis que sin embargo sugiere la necesidad de algunas revisiones profundas en nuestra comprensión de la naturaleza de las representaciones internas y de los modelos interiores del mundo. Los retos de cimentación y conceptuales se muestran así bastante reales, aunque privados de sus radicales adornos postcartesianos.

## II. LA HUIDA DEL SÍMBOLO INTERIOR.

El rechazo total de la noción de representación interna es justamente el caso límite extremo de una notable tendencia que podría ser calificada como «la huida del símbolo interior». Esta huida conlleva el rechazo paulatinamente progresivo del dispositivo y las nociones asociadas con la visión de la cognición como la manipulación de símbolos interio-

6 Las mejores exposiciones de este punto de vista están en J. Haugeland, "Mind Embodied and Embedded", en Y.-H. Houng y J.-C. Ho (eds.), *Mind and Cognition* (Taipei, Taiwan: Academia Sinica, 1995), pp. 3-38, y T. Van Gelder, "What Might Cognition Be, If Not Computation?" *Journal of Philosophy* 92/7 (1995), 345-81. Hay afirmaciones y argumentos muy cercanamente relacionados en T. Van Gelder y R. Port "It's About Time" Introducción a R. Port y T. Van Gelder (eds.), *Mind as Motion: Dynamics, Behavior, and Cognition* (Cambridge, MA: MIT Press, 1995). E. Thelen y L. Smith, *A Dynamic Systems Approach to the Development of Cognition and Action* (Cambridge, MA: MIT Press, 1994), y Varela, Thompson y Rosch, *The Embodied Mind*.

7 Este punto de vista está claramente contemplado en Haugeland "Mind Embodied" y en Van Gelder. "What Might Cognition Be?". Ambos autores, no obstante, reconocen el amplio margen de posibilidades intermedias. El término 'agente postcartesiano' es de Van Gelder, p. 381. Ver también Thelen y Smith, *A Dynamic Systems Approach*, p. 338, y Van Gelder y Port, "It's About Time", p. ix.

res troceados. De acuerdo con una visión simple (e históricamente importante), las transiciones semánticamente sensibles entre estados mentales están mejor explicadas en términos de transiciones sintácticamente forzadas entre series simbólicas interiores. Estas series simbólicas contenían elementos discretos que se corresponderían bastante más cercanamente con los elementos semánticos identificados en las descripciones oracionales de los estados mentales relevantes. Así pues, el pensamiento de que Juan ama a María se imagina que se ejecuta como una compleja serie simbólica interior que incorpora elementos distintos e independientemente manipulables representando a 'Juan', 'ama' y 'María'<sup>8</sup>.

Esta visión de simples átomos simbólicos interiores (elementos básicos desestructurados) que se corresponde bastante cercanamente a los conceptos familiares y relaciones envueltos en el discurso diario fue impugnada por el desarrollo de los modelos distribuidos conexionistas<sup>9</sup>. El paradigma oracional<sup>10</sup> fue sustituido, en esta investigación, por una visión de las representaciones internas como patrones de actividad distribuidos a lo largo de una completa formación de unidades de procesamientos simples. A tales patrones distribuidos se les permitió solaparse con modos semánticamente significativos, dando lugar a una variedad de efectos colaterales computacionalmente significativos, incluyendo la generalización libre, la resistencia al deterioro, etc.<sup>11</sup>.

Más recientemente aún, hemos sido testigos de la creciente atención a la dinámica temporal de los vehículos internos de representación. El uso de (por ejemplo) sencillas redes neuronales recurrentes<sup>12</sup> permite que la información sea codificada no exactamente en patrones de actividad instantáneos sino en trayectorias de procesamiento temporalmente extendidas. En estas redes, la mayor parte del poder para procesar información reside en el modo como un estado presente permite o restringe el cambio y la evolución futuros. La progresión ha sido entonces desde

8 Ver J. Fodor y Z. Pylyshyn, "Connectionism and Cognitive Architecture: A critical analysis", *Cognition* 28, (1988), p. 13.

9 Ver D. Rumelhart, J. McClelland y el grupo de investigación PDP, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vols. I y II (Cambridge, MA: MIT Press, 1986).

10 P. S. Churchland, *Neurophilosophy* (Cambridge, MA: MIT Press, 1986).

11 Los detalles no nos afectan ahora. Ver, no obstante, por ejemplo mi *Microcognition: Philosophy, Cognitive Science and Parallel Distributed Processing* (Cambridge, MA: MIT Press, 1989) para una discusión.

12 J. Elman, "Representation and Structure in Connectionist Models", en G. Altman (ed.), *Cognitive Models of Speech Processing* (Cambridge, MA: MIT Press, 1991).

una consideración de los simples símbolos atómicos internos hacia una noción de patrones espacialmente distribuidos y hacia una noción de patrones distribuidos espacial y *temporalmente*. Los vehículos internos de contenido, cortesía de esta progresión, vienen a ser cada vez menos como *estados* internos y cada vez más como *procesos* internos crecientemente complejos.

Esta metamorfosis, además, está probablemente incompleta aún. Los desarrollos futuros procurarán incluir la comprensión de muchos vehículos internos como polivalentemente funcionales y los aspectos comprensivos de la arquitectura interna como dinámicamente reconfigurables. La funcionalidad múltiple significaría que uno y el mismo recurso puede jugar una variedad de papeles portadores de contenido<sup>13</sup> (quizás variando en función del contexto local). La reconfigurabilidad dinámica significaría que la arquitectura interna está ella misma sujeta a rápido cambio y reorganización, como cuando la liberación de un neuro-modulador químico da lugar a dos redes neuronales para fundirlas temporalmente y comportarse como una.

La moraleja es, pues, que nuestra comprensión de la naturaleza de los (putativos) vehículos internos de contenido está en un estado de extremado cambio, caracterizado por una rápida huida desde la inicial imagen de estáticos y troceados símbolos internos desestructurados. Esta huida tiene también un aspecto relacionado de contenido. Así como los vehículos internos se han hecho más complejos, los contenidos característicos parecen haberse hecho más parciales y fragmentarios. Ello es así porque el énfasis se ha trasladado desde las formas aislacionistas de resolución de problemas hacia series reiteradas de interacciones agente-entorno. Este traslado está magistralmente ejemplificado en el trabajo reciente en el campo conocido como Visión Animada<sup>14</sup>.

Recuérdese la descripción de Marr<sup>15</sup> de la tarea de la visión. Esta tarea, según Marr, es la de construir un rico modelo interior de la escena visual tridimensional sobre la base de la información de entrada disponible (bidimensional). El trabajo reciente en el campo conocido como Visión Animada ha tomado un rumbo muy diferente. Este trabajo describe la tarea como, simplemente, el uso de estrategias visuales para controlar la conducta, en contextos del mundo real, al menor coste

13 Para algunas sugerencias sobre tal complejidad sensible al contenido, ver J. Knierim y D. Van Essen, "Virtual Cortex: Cartography, Connectivity and Concurrent Processing", *Current Opinion in Neurobiology* 2 (1992), 150-5.

14 D. Ballard, "Animate Vision", *Artificial Intelligence* 48 (1991), 57-86.

15 D. Marr, *Vision* (San Francisco, CA: W. H. Freeman, 1982).

computacional como sea posible. Para estos efectos la Visión Animada dispone por sí misma de tres estratagemas centrales

1. El uso de claves de tarea específica y atajos.
2. El uso de estrategias centradas en el cuerpo (egocéntricas).
3. El uso de interacciones repetidas con el entorno.

Las claves de tarea específica y los atajos incluyen, por ejemplo, el uso de estrategias idiosincrásicas personalizadas tales como la búsqueda del amarillo brillante (una clave visual barata y fácil) cuando busco *mi* taza de café (que acontece exactamente que es de un amarillo chillón). Las estrategias egocéntricas incluyen el uso de los llamados indicadores deícticos (explicados luego). Las interacciones repetidas con el entorno incluyen, por ejemplo, el uso de repetidos movimientos rápidos del ojo para visitar y revisitarse diferentes aspectos de una escena recuperando la información específica sólo como y cuando se requiere.

El caso de los indicadores deícticos puede servir como ilustración general. Un indicador es en Inteligencia Artificial clásica un estado interior que puede funcionar en rutinas computacionales autocontenidas pero que también puede señalar a otras estructuras de datos<sup>16</sup>. Este señalamiento permite la recuperación de una información más detallada, cuando ello sea requerido, y la ligazón efectiva de ciertos componentes de información a otros. Tal ligazón necesita ser a menudo temporal, como cuando ligamos ciertas características (V.g.: el amarillo brillante) a ciertas localizaciones visuales actuales, aunque claramente temporales (V.g.: «el amarillo detectado en la parte alta de la izquierda del campo visual»).

Los indicadores *deícticos*, no obstante, son orientaciones corporales reales (tales como los movimientos rápidos del ojo) que juegan la misma clase de papel funcional. La idea es que el sistema está establecido de tal manera que el mismo acto de fijar un aspecto particular de una escena visual implementa una clase de ligazón temporal variable en la que las características detectadas están ligadas a una ubicación espacial dada. Un ejemplo cercano afecta al modo de ligar una rutina de «alcance y agarre» al objeto que sirve de objetivo. Aquí también la ligazón puede ser implementada de forma económica usando lo que se llama informalmente la estrategia del «hazlo donde estoy mirando». Aquí el sistema está establecido de tal manera que la rutina de agarre está automática-

16 Ver, por ejemplo, Z. Pylyshyn (ed.), *The Robot's Dilemma: The Frame Problem in Artificial Intelligence* (Norwood: Ablex, 1987).

mente dirigida a la localización visual fijada actualmente. En todos estos casos los autores comentan:

El mundo externo es análogo a la memoria de un ordenador. Cuando fijamos una localización, las neuronas que están vinculadas a la fovea se refieren a la información computada desde esa localización. El cambio de mirada es análogo al cambio en la referencia de la memoria en un ordenador de silicio<sup>17</sup>.

Un avance importante en la investigación de la Visión Animada consiste, pues, en que las acciones corporales (tales como los movimientos rápidos del ojo) pueden jugar papeles computacionales vitales. Otro avance consiste en que la interacción repetida agente-entorno evita en gran parte la necesidad de crear detallados modelos internos del mundo de propósito general. Al visitar y visitar diferentes aspectos de la escena actual como y cuando se requiera, permitimos que el mundo funcione como «su propio mejor modelo». El programa de investigación es, entonces, fielmente antiaislacionista. Pero no es en modo alguno «postcartesiano» –no rechaza precisamente tanto las ideas de modelos y representaciones internos, como que las reconfigura en una imagen más dispersa y más interactiva. Leemos aún sobre «las bases de datos internos» (V.g.: las que asocian objetos pequeños, como las llaves de mi coche, con localizaciones más grandes, fácilmente detectables, como que están sobre la mesa de la cocina), sobre «representaciones internas características» (de color, forma, etc.), sobre «representaciones indexicales» que especifican las localizaciones relativas a la posición corporal y cosas así. Lo que está siendo rechazado enfáticamente *no* es la noción de estados internos portadores de contenidos per se, sino más bien la noción mucho más específica de que construimos ricas representaciones internas de memoria profunda de todos los aspectos de la escena visual presente.

Un perfil similar se presenta en la investigación más actual sobre la robótica en el mundo real. Un buen ejemplo es el trabajo<sup>18</sup> en la nave-

17 D. Ballard, M. Hayhoe, P. Pook y R. Rao, “Deictic Codes for the Embodiment of Cognition”, *Behavioral and Brain Sciences*, de próxima aparición.

18 M. Matricic, “Navigating with a Rat Brain: A Neurobiologically Inspired Model for Robot Spatial Representation”, en J.-A. Meyer y S. Wilson (eds.), *From Animals to Animals 1* (Cambridge, MA: MIT Press, 1991). Este trabajo se discute ampliamente en Clark, *Being There*, capítulo 2.

gación robótica del mundo real en la que el conocimiento de la localización está directamente codificado como una rutina perceptomotora: una rutina que especifica realmente cómo trasladar al robot desde su posición presente a la localización de destino. En estos modelos el mapa interior es *él mismo* el controlador de la acción apropiada. No hay necesidad alguna de un sistema ulterior para acceder al mapa y planificar una ruta. El conocimiento del robot es entonces descriptivo y prescriptivo a la vez<sup>19</sup> —una naturaleza dual que permite grandes economías en términos de tiempo de respuesta y esfuerzo computacional. De nuevo vemos un interesante trabajo que no es tanto antirepresentacional cuanto dispersamente representacional. La distinción crucial, tengo para mí, no está tanto entre las soluciones representacionales y las no representacionales cuanto entre las formas ricas y caras de representación interna (que pueden incrementar la flexibilidad pero que requieren a menudo un trabajo computacional adicional para especificar una respuesta conductual) y las más dispersas formas de representación más orientadas a la acción.

El trabajo más convincente en Visión Animada y robótica del mundo real se para entonces muy cerca del entero rechazo «postcartesiano» de los modelos y las representaciones internos ¿Por qué, pues, algunos teóricos han continuado cuestionando la idea de representaciones interiores y modelos internos *tout court*?

### III. INTERACCIONISMO RADICAL

El argumento<sup>20</sup> antirepresentacionista principal parece disparar el impacto de los densos intercambios causales recíprocos que unen agente y entorno en una compleja red de influencias mutuas. Bajo tales condiciones, se argumenta, las clases de descomposición representacional y el análisis que funciona tan bien para muchos modelos contemporáneos de ordenador de procesos inteligentes simplemente no hacen pie. El problema (se arguye) es que la noción de *x* representando a *y* es demasiado unidireccional y demasiado simplista para hacer justicia a casos en los que *x* está afectada y es afectada continuamente por *y*, y viceversa. Sin embargo las típicas interacciones agente-entorno, continúa el argumento, presentan a menudo exactamente tal perfil causal complejo y circular.

19 Para más información sobre este asunto, ver R. Millikan "Pushmi-pullyu Representations", en L. May; M. Friedman y A. Clark (eds.), *Mind and Morals* (Cambridge, MA: MIT Press, 1996).

Consideremos un baile de salón. En la medida en que bailas, tus movimientos (¡si eres un buen bailarín!) están continuamente influidos por, e influyen a, los de tu pareja: los dos juegos de movimientos «coevolucionan» de un modo grandemente interdeterminado. Y no es esencial para el fenómeno la presencia de dos agentes humanos. La misma relación compleja se consigue entre (por ejemplo) un windsurfista experimentado y su aparejo: el windsurfista influye y es influido constantemente por el conjunto del aparejo. Van Gelder aclara lo mismo usando el conocido ejemplo del regulador de Watt (o centrifugo) —un instrumento que mantiene una máquina de vapor a una velocidad constante de dos maneras: afectando y siendo afectado por la velocidad de la máquina<sup>21</sup>. Tales episodios de mutua influencia fueron muy discutidos por la temprana cibernética<sup>22</sup> y en la obra del fenomenólogo francés Maurice Merleau-Ponty<sup>23</sup>.

Donde tal influencia causal, continua, densa y circular se consigue, se argumenta, las herramientas del análisis representacional (y computacional) están embarrancados. La idea de explicar la forma de esas complejas interacciones en curso agente-entorno mediante la descripción de los estados internos como representando los externos se rechaza como vulgar y no esclarecedora. En cambio, lo interior y lo exterior coevolucionan de un modo matemáticamente preciso que se capta mejor (así sigue el argumento) mediante el uso de ecuaciones diferen-

20 Este argumento es la piedra angular de “What Might Cognition Be?”, de Van Gelder, donde leemos, por ejemplo, que: “La hipótesis nuclear dinámica... va de la mano de una concepción de los sistemas cognitivos... como complejos de un cambio continuo, simultáneo y mutuamente determinante. [] En esta visión el sistema cognitivo no es exactamente el cerebro encapsulado; sino que, más bien, puesto que el sistema nervioso, el cuerpo y el entorno están todos cambiando constantemente e influenciándose unos a otros simultáneamente, el verdadero sistema cognitivo es un único sistema unificado que abraza a los tres” (p. 375). El argumento es también visible en Van Gelder y Port. “It’s About Time”, pp. 23-5; en Thelen y Smith, *A Dynamic Systems Approach*, p. 27; y en Varela, Thompson y Rosch, *The Embodied Mind*, pp. 172-5.

21 Aquí Van Gelder, “What Might Cognition Be?” (p. 353), apunta que: “el ángulo del brazo y la velocidad de la máquina están todo el tiempo determinados por, y determinando, cada uno la conducta del otro... no hay nada arcano en esta relación... Sin embargo, es mucho más sutil y complejo de lo que el concepto habitual de representación puede dar de sí.” Este ejemplo se trata detalladamente en A. Clark y J. Toribio, “Doing Without Representing?” *Syntese* 101 (1995), 401-31.

22 Por ejemplo, en W. Ross Ashby, *Introduction to Cybernetics* (Nueva York, Wiley, 1956).

23 Maurice Merleau-Ponty, *The Structure of Behavior* (New York: Beacon, 1963). Originalmente *La Structure du Comportement* (Paris: Presses Universitaires de France, 1942).

ciales acopladas en las que los valores actuales de ciertas variables internas aparecen como engarces de parámetro en la ecuación de evolución para el sistema externo y viceversa<sup>24</sup>. Afortunadamente los detalles de tal modelo de sistemas dinámicos no son relevantes para los propósitos presentes<sup>25</sup>. Lo que importa es la forma general del argumento. Van Gelder lo puntualiza bien:

La operación interna de un sistema que interactúa con un mundo externo puede ser tan sutil y compleja como para desafiar la descripción en términos representacionales. (p. 381).

Antes de responder a este argumento conviene hacer una pausa para clarificar el reto. Pues lo que está en cuestión no es el estado de ciertos sistemas (nosotros mismos, por ejemplo) como *representadores*. Eso es un dato. Desde luego que nos representamos nuestro mundo, nuestro pasado, nuestros posibles futuros, nuestros amigos ausentes y cosas así. Pensamos en esas cosas y estados de cosas y hasta cierto punto las representamos claramente para nosotros mismos. Lo que *no* es un dato (y que es lo que está aquí sobre el tapete) es que nosotros nos valemos de *representaciones internas* para hacer eso. El asunto es que la pretensión científica de que la cognición envuelve representaciones internas (y computaciones definidas sobre ellas) no es tenida como un simple ensayo del hecho de que somos pensadores, sino como una hipótesis empírica sustancial y explicativamente potente: la clase de cosa que desde luego podría resultar falsa. La pretensión falsable, en una primera aproximación, es que hay distintos estados o procesos interiores identificables, cuyo papel sistemático o funcional está para sustituir características específicas o estados de cosas.

Esta noción de sustitutos internos es, no obstante, ambigua por sí misma. Es ambigua<sup>26</sup> entre una noción débil en la que  $x$  «sustituye» a  $y$  y  $x$  es un recurso interno que (a) transporta información sobre  $y$ , y (b) se usa para controlar la conducta, y una noción mucho más fuerte en la que el recurso interno debe ser capaz de funcionar como un sustituto auténtico, esto es, que sea capaz de controlar sistemáticamente la conducta apropiada, *incluso si y está ausente o es inexistente*. Una pobla-

24 Para una introducción accesible a estos planteamientos dinámicos, ver S. Kelso *Dynamic Patterns* (Cambridge, MA: MIT Press, 1995). Un texto clásico es el de R. Abraham y S. Shaw *Dynamics - The Geometry of Behavior* (Redwood, CA: Addison-Wesley, 1992).

25 Para una discusión más completa, ver Clark, *Being There*, capítulos 5, 6 y 8.

26 Ver A. Clark y R. Grush, "Towards a Cognitive Robotics" (remitido para ser publicado).

ción neuronal<sup>27</sup> estrechamente acondicionada para la orientación corporal y usada para controlar la acción especializada en línea puede entonces considerarse como un sistema de sustitutos débiles. E incluso aquí la glosa representacional parece decirnos algo útil sobre el objetivo de la población neuronal. Pero tal población, aunque se compromete en el control de la acción basada en la información, no tiene porqué ser capaz de conducir las acciones apropiadas en ausencia del estado de cosas (débilmente representado). Es esta segunda, y ciertamente menos común, capacidad de obrar como un sustituto interno *en ausencia de un directo control ambiental* la que, sugiero yo, caracteriza los casos más fuertes y conceptualmente inequívocos de representación interna<sup>28</sup>.

El problema es pues que el argumento entero sobre la complejidad circular y causal de ricas interacciones agente-entorno está viciado por su fallo en comprometer el resultado real de una representación fuerte. Todos los ejemplos comparten (y deben compartir) una cierta característica problemática, a saber, que todos ellos son casos en los que la conducta final *está continuamente dirigida y modificada por* el parámetro ambiental relevante. Sin embargo una mayor motivación para proponer las representaciones internas en primer lugar fue la de explicar nuestra misteriosa capacidad para traspasar límites como las fortísimas interacciones acopladas agente-mundo y coordinar nuestras actividades y elecciones con lo distal, lo modal y lo no existente. La noción original de representación interna está entonces fundada en la noción de fuertes sustitutos internos y se extiende meramente (quizás problemáticamente) al caso de (meros) estados internos portadores de información usados para el control de la acción. Esto ayuda a explicar por qué los mejores casos del argumento-desde-la-continua-causación-recíproca pueden antojársenos ejemplos bastante *pobres* de los fenómenos tradicionalmente cognitivos. Ya que dependen crucialmente de la presencia constante de factores ambientales relevantes y entonces no se nos antojan especialmente necesitados de representación<sup>29</sup> en primer lugar. Los escenarios necesitados de representación deberían ser propiamente, por ejemplo, la

27 Por ejemplo, la población posterior parietal neuronal de una rata que codifica la información sobre el modo como la cabeza de la rata está orientada y que se explota en el recorrido del laberinto radial. Ver B. Naughton y L. Nadel "Hebb-Marr Networks and the Neurobiological Representation of Action in Space", en M. Gluck y D. Rumelhart (eds.), *Neuroscience and Connectionist Theory* (Erlbaum, 1990).

28 David Israel "Bogdan on Information" *Mind & Language* 3/2 (1988), 123-40 hace esencialmente la misma observación. Ver también Brian Cantwell Smith, *The Origin of Objects* (Cambridge, MA: MIT Press, 1996).

29 La frase es de Clark y Toribio, "Doing Without Representing?".

planificación de las vacaciones del año próximo, usar imágenes mentales para contar el número de ventanas de tu vieja casa (este ejemplo me lo contó Dan Dennett), hacer aritmética mental, soñar, etc., etc.

La situación dialéctica es, no obstante, bastante delicada. Pues el antirepresentacionista puede ahora replicar que el meollo de su argumento fue, en parte, sugerir que estos casos tradicionales (de la que podría ser llamada razón «ambientalmente desacoplada») son de hecho empíricamente marginales y que la masa de la respuesta inteligente cotidiana muestra *precisamente* el perfil ricamente interactivo que ilumina el argumento. La razón ambientalmente desacoplada, se dice, es en el mejor de los casos un fenómeno de la punta del iceberg. Lo que se está promocionando entonces es un cambio de énfasis desde la cogitación fuera de línea hacia el compromiso<sup>30</sup> interactivo en tiempo real –una clase de cambio de lo cognitivo a lo ejecutivo.

Este cambio de énfasis es bienvenido en un sentido. Desde un punto de vista evolutivo y de desarrollo<sup>31</sup>, la sensibilidad al mundo real y al tiempo real es claramente primaria en algún sentido. Pero, como veremos ahora, la noción de que el caso ricamente interactivo es de algún modo *biológicamente básico* es de hecho perfectamente compatible con la tesis de que la razón ambientalmente desacoplada fuera de línea *no* es la mera punta del iceberg adaptativo. Desde luego, el camino para forjar una auténtica ciencia cognitiva de la acción incorporizada y ambientalmente incorporada está, según creo, precisamente para ser el destino de las relaciones entre lo densamente acoplado y las formas más fuerte y representacionalmente mediatas del éxito adaptativo. Éste es el proyecto que yo llamo Cartesianismo Mínimo y hacia el que vamos a volver ahora.

#### IV. CARTESIANISMO MÍNIMO

El cartesianismo mínimo ambiciona localizar las raíces de la razón fuertemente representacional en los engarces ricamente interactivos resaltados en el trabajo sobre la cognición incorporizada. De esta manera considera el fenómeno del alcance especializado<sup>32</sup>. El alcance especia-

30 Este paso está explícitamente dado en Haugeland, “Mind Embodied” y es claramente evidente en Van Gelder y Port, “It’s About Time”.

31 Ver Thelen y Smith, *A Dynamic Systems Approach*.

32 Tomo prestado este caso de R. Grush, “Emulation and Cognition” (Tesis doctoral, University of California at San Diego). Un tratamiento ulterior se encuentra en Clark y Grush, “Towards a Cognitive Robotics”.

lizado suave implica el uso de retroalimentación propioceptiva —señales que dicen al cerebro cómo está orientado el brazo en el espacio. Pero el cronometraje de esas señales plantea un problema. El retraso mínimo entre el comienzo y el uso de tal información parece estar entre 200 y 500 milisegundos<sup>33</sup>. Sin embargo hacemos correcciones de la trayectoria esencial, que parecen estar gobernadas por tal retroalimentación, dentro de los primeros 70 milisegundos<sup>34</sup> de alcance. ¿Cómo resuelve el problema la naturaleza?

Este problema de requerir retroalimentación antes de que esté prácticamente disponible aparece también en la industria: en las plantas químicas, reactivos biológicos y demás<sup>35</sup>. En estos casos una solución corriente es la de añadir un modelo adelantado o emulador a los sistemas. Éste es un circuito que toma como entrada una especificación de dos cosas, del estado previo del sistema y de las órdenes recién emitidas, y que da como salida una *predicción* de la retroalimentación que debería llegar más tarde. Entonces el emulador genera una especie de señal de retroalimentación simulada disponible sustancialmente por adelantado a la cosa real.

Parece ahora que la naturaleza puede desplegar la misma estrategia. Hay un cuerpo creciente de pruebas<sup>36</sup> neurocientíficas que sugieren que la circuitería neuronal que abarca el tracto corticoespinal, el núcleo rojo, la oliva inferior, el cortex contralateral dentado y el cerebelar pueden estar jugando exactamente tal papel. Esta circuitería parece hacer una copia de la orden motora aferente y producir una predicción rápida de la retroalimentación que debe llegar más tarde por la ruta lenta de los 200-500 milisegundos.

33 Esta cifra se halla, por ejemplo, usando vibradores artificiales atados a los tendones para interrumpir las señales propioceptivas que llegan de los husos del músculo y calcular la duración del intervalo entre esa entrada perjudicial y las alteraciones en el movimiento mismo del brazo (ver C. Redon, L. Hay y J. L. Velay, "Proprioceptive Control of Goal Directed Movements in Man, Studied by Means of Vibratory Muscle Tendon Stimulation" *Journal of Motor Control* 23/2 (1991), 101-8).

34 Ver J. Van der Meulen, R. Gooskens, J. J. Dennier van der Gon, C. C. A. M. Gielen y K. Wilhelm, "Mechanisms Underlying Accuracy in Fast Goal-directed Arm Movements in Man", *Journal of Motor Behavior* 22/1 (1990), 67-84.

35 Para una revisión ver Grush, "Emulation and Cognition".

36 Ver M. Ito, *The Cerebellum and Neural Control* (Nueva York: Raven Press, 1984); M. Kawato, K. Furukawa y R. Suzuki, "A Hierarchical Neural Network Model for the Control and Learning of Voluntary Movement", *Biological Cybernetics* 57 (1987), 169-85; y D. Wolpert, Z. Ghahramani y M. Jordan, "An Internal Model for Sensorimotor Integration", *Science* 269 (1995), 1880-2.

El mismo truco se ha reproducido en una diversidad de modelos de redes<sup>37</sup> neuronales. Lo que importa para nuestros propósitos es, no obstante, una conjetura adicional. Esta conjetura<sup>38</sup> consiste en que el circuito del emulador biológico juega un doble papel. Este papel dual conlleva dos cosas, la buena sintonía del alcance en línea (el caso normal en que el circuito del emulador actúa como ayuda para nivelar el alcance en tiempo real) y la producción de imágenes viso-motoras que permiten *el ensayo mental fuera de línea* de las rutinas motoras. En el último caso el circuito emulador está ahora funcionando solo, desacoplado del sistema de acción en el mundo real. Tal papel adicional para la misma circuitería emuladora implicada en el alcance especializado diario parece plausible evolutivamente y ayuda a explicar de alguna otra forma los resultados misteriosos. Éstos incluyen el resultado fuerte de que el ensayo mental puede mejorar realmente las habilidades deportivas<sup>39</sup> y la actividad sorprendente del cerebelo (generalmente pensado como una zona motora) durante las imágenes mentales<sup>40</sup>.

La circuitería de emulación motora, si esto es correcto, es dos cosas, una ayuda para la acción fluida en el mundo real y un apoyo para el ensayo mental independiente y desacoplado ambientalmente. Entonces es una herramienta mental mínimamente cartesiana, pero una herramienta que es parasitaria de adaptaciones cercanamente ajustadas a la promoción de las suaves interacciones en tiempo real agente-entorno. Como resultado, incluso el fenómeno modestamente cartesiano de la imaginación visual permanece cercanamente unido a la biomecánica y al perfil generador de acción del agente.

Dado este perfil podemos ver por qué las metodologías y las suposiciones aislacionistas pueden mostrarse inadecuadas incluso en el caso de ciertas clases de habilidades cognitivas ambientalmente desacopladas. Pues tales habilidades pueden permanecer orientadas a la acción relacionada con ellas, gracias a los constreñimientos sobre la dotación original: una dotación que es ahora redesplegada para servir a fines

37 V.g.: M. Kawato, "Computational Schemes and Neural Network Models for Formation and Control of Multijoint Arm Trajectory", en W. T. Miller III, R. Sutton y P. Werbos (eds.), *Neural Networks for Control* (Cambridge, MA: MIT Press, 1990); y Wolpert *et al.*, "An Internal Model".

38 Grush, "Emulation and Cognition".

39 Ver, por ejemplo, D. Fetz y D. Landers, "The Effects of Mental Practice on Motor Skill Learning and Performance: A Meta-Analysis", *Journal of Sport Psychology* 5 (1983), 25-57.

40 J. Decety, H. Sjöholm, E. Ryding, G. Stenberg y D. Ingvar, "The Cerebellum Participates in Cognitive Activity", *Brain Research* 535 (1990), 313-17.

«cartesianos fuera de línea». Decisivamente, este fracaso del aislacionismo no debería ser visto como una invitación al escepticismo sobre la representación interna y los modelos interiores. En el caso del emulador, al menos, es claramente evidente que estamos ahora lidiando con una circuitería identificable cuyo papel funcional es, a veces, fuertemente representacional. Sin embargo es perfectamente compatible la explicación con diversas moralejas y subrayados sugeridos por la investigación orientada a la acción previamente discutida en la sección II. La posición conciliadora que yo apoyo conlleva entonces combinar el énfasis sobre la acción en el mundo real y en tiempo real con una búsqueda de las raíces biológicamente básicas de formas más desacopladas de pensamiento y resolución de problemas. Pues es solamente con la confrontación de la última clase de casos con lo que el representacionismo puede tener un juicio justo.

#### V. INCREMENTO, RACIONALIDAD Y COMPLEJIDAD.

El cartesianismo mínimo busca tender puentes entre la reciente insistencia en las tareas ricamente interactivas y el enfoque cognitivo más tradicional sobre la razón desacoplada. Para ese fin acentúa el uso de múltiples modelos internos parciales orientados a la acción y de internas representaciones deícticas, idiosincrásicas y orientadas a la acción. La cuestión obligada en este punto se convierte en si realmente podemos tener esperanzas en explicar la serie entera de la cognición humana sin terminar por reinventar en algún momento la imagen clásica de las estructuras simbólicas interiores, neutras al contexto, ricas, independientes de la acción y altamente manipulables. En resumen, ¿puede el cartesianismo mínimo incrementarse tanto como para explicar las enteras complejidades de la «cognición más superior»?

Tal «incremento», si tiene que tener una oportunidad razonable de éxito, debe dar el debido crédito al modo en que las estructuras externas, las acciones lingüísticas y las prácticas culturales conspiran todas para reconfigurar efectivamente la forma de los espacios computacionales que debemos salvar de cara a resolver problemas más complejos y abstractos. La compleja cognición humana se describe mejor como sucediendo en la fecunda interfaz entre una variedad de recursos internos orientados a la acción y un amplio andamio de estructuras, herramientas y prácticas externas: una red de seguridad que lo mismo actúa para alterar sustancialmente espacios computacionales que puede ser explorada por nuestra forma de razón biológica básica e incorporada. Un ejemplo

clásico<sup>41</sup> es el uso de la pluma y el papel para hacer (por ejemplo) una multiplicación larga: un truco que nos permite usar una secuencia repetida de cálculos internos simples (tales como  $7 \times 7$ ,  $4 \times 4$ ) y una secuencia de inscripciones externamente almacenadas y manipuladas que permiten resolver problemas mucho más complejos (tales como  $777 \times 444$ ). El lenguaje público, arguyo en otra parte, juega una amplia variedad de papeles similares<sup>42</sup>. El mero acto de etiquetar, como Dennett<sup>43</sup> sugiere, permite grandes economías de búsqueda y clasificación, mientras que la capacidad de ensayo lingüístico puede ser, según Ray Jackendoff<sup>44</sup>, la que nos permite ocuparnos de los detalles de nuestros propios pensamientos –abriendo así vastas posibilidades nuevas de reflexión y análisis<sup>45</sup>. Los artefactos externos y las organizaciones sociales alteran y transforman asimismo las tareas que los cerebros individuales necesitan ejecutar. El antropólogo cognitivo Ed Hutchins<sup>46</sup> ofrece a este respecto un relato maravillosamente detallado y convincente del proceso de la navegación marítima en el que el sistema entero consta de múltiples cerebros, cuerpos e instrumentos que resuelven el problema de la navegación. Cada miembro de la tripulación dentro de este nexo más amplio meramente controla y responde a ciertas condiciones ambientales simples. Las respuestas alteran unos pocos aspectos del espacio de trabajo compartido y así promueven y mantienen formas similares de reacción entre los otros. El proceso entero constituye un flujo computacional ambientalmente extendido en el que agentes múltiples, rutinas simples y una variedad de accesorios externos y artefactos (tales como las reglas de cálculo náutico) se combinan para resolver un problema complejo.

Incluso un tratamiento mínimamente cartesiano de la razón biológica básica puede de este modo tener la esperanza de incrementar lo sufi-

41 D. Rumelhart, J. McClelland, P. Smolensky y G. Hinton, "Schemata and Sequential Thought Processes in PDP Models", en D. Rumelhart, J. McClelland and the PDP Research Group (eds.), *Parallel Distributed Processing: Explanations in the Microstructure of Cognition*, Vol. II (Cambridge, MA: MIT Press, 1986), pp. 7-58.

42 A. Clark, "Magic Words: How Language Augments Human Cognition", en P. Carruthers y J. Boucher (eds.), *Thought and Language: Interdisciplinary Themes* (Cambridge: Cambridge University Press, de próxima aparición).

43 D. Dennett, "Labelling and Learning", *Mind and Language* 8 (1994), 540-8. Ver también el capítulo 13 de su *Darwin's Dangerous Idea* (Nueva York: Simon & Schuster, 1995).

44 R. Jackendoff, "How Language Helps Us Think", *Pragmatics and Cognition* 4/1 (1996), pp. 1-34.

45 Ibid., pp. 19-22.

46 E. Hutchins, *Cognition in the Wild* (Cambridge, MA: MIT Press, 1995).

ciente como para iluminar la entera panoplia del pensamiento y razón humanos. El truco está en tomar de hecho la cuestión del andamio externo muy seriamente (y especialmente reconocer las virtudes computacionales del lenguaje público: el único código simbólico neutral a la acción que ya sabemos que poseemos nosotros mismos). Una implicación de este enfoque del problema del incremento es que necesitaremos, a veces, estudiar esos sistemas más amplios (de múltiples cerebros y artefactos en comunicación) como conjuntos organizados y reconocer los procesos computacionales ampliados abarcando los linderos entre el cerebro, el cuerpo y el mundo. Tales afirmaciones pueden ser fácilmente malinterpretadas por antipatía hacia el estudio de los recursos y procesos internos. Pero el reto real, una vez más, es entrelazar las dos aproximaciones y así reubicar a la razón humana individual en su propio nicho ecológico.

El proyecto también plantea cuestiones sobre la misma noción de racionalidad humana. La ciencia cognitiva aislacionista tiende a describir la racionalidad en términos de transiciones semánticamente aptas entre estados mentales internos. El acierto de Turing, como repetidamente recalcó Jerry Fodor<sup>47</sup>, fue mostrar cómo tales transiciones podrían ser mantenidas por un proceso puramente mecánico. El enfoque ambientalmente ampliado ahora debatido no necesita rechazar esa descripción. Éste puede (y debería) incorporar la idea central de Turing sobre procesos interiores cuyas propiedades sintácticas<sup>48</sup> preservan relaciones semánticas. Pero esto será justamente parte de una teoría más abarcadora que permita a la conducta racional sobrevenir a las más amplias redes de una estructura que envuelve a otros agentes, artefactos y aspectos del entorno local.

Queda un desasosiego sobre la complejidad. Incluso si el proyecto general bosquejado en este trabajo se muestra atractivo (el proyecto de construir un puente entre los modelos basados en la interacción y las formas de razón más ambientalmente desacopladas), podría aún resultar que los vehículos internos de contenido se muestren demasiado espacial y temporalmente complejos como para representar descripciones

47 Ver, por ejemplo, los comentarios en las páginas 277-8 y su "Replies to Critics", en B. Loewer y G. Rey (eds.), *Meaning in Mind: Fodor and his critics* (Oxford: Blackwell, 1991), pp. 255-319.

48 Las propiedades sintácticas son algunas propiedades no semánticas que pueden ser explotadas directamente por un sistema físico. Los procesos ampliados temporalmente son también sintácticos en este sentido, tal y como se describen en la sección II.

iluminadoras de los procesos mentales. Tal desasosiego gana alguna fuerza desde las recientes demostraciones del papel de las complejas conexiones<sup>49</sup> recurrentes para modular el perfil del procesamiento de información de poblaciones neuronales y desde la escabrosa dificultad para asignar papeles específicos portadores de contenido a los rastros de la maquinaria neuronal. Estas complejidades y dificultades pueden llevar a una clase de escepticismo sutilmente diferente en el que es la complejidad del relato mismo (no la interacción interior-exterior) la que se supone plantea dificultades para el análisis representacional.

Aquí las cuestiones son más puramente empíricas y es imposible, dado el estado actual de la investigación, hacer predicciones firmes. Pero una posibilidad interesante es que nuevas herramientas analíticas puedan aún proporcionar los medios para identificar patrones de actividad funcionalmente importantes. Los análisis de los sistemas dinámicos, de la clase a veces fomentada como una alternativa al enfoque representacional, pueden de hecho ayudarnos a identificar vehículos internos controlables a pesar de la presencia de la complejidad espacial y temporal que rebrota. Esta posibilidad está claramente señalada por el mismo van Gelder<sup>50</sup>, quien concede que «una característica apasionante del enfoque dinámico es que ofrece oportunidades para volver a imaginar notablemente la naturaleza de la representación en los sistemas cognitivos». Las representaciones internas, pues, pueden ser comprendidas no como simples estados internos sino como patrones dinámicos de casi cualquier clase concebible. Tales patrones pueden, además, ser entidades transitorias que se forman solamente como respuesta a los detalles del contexto presente. Entonces apreciamos mejor los límites de la misma metáfora del vehículo interior. Tales vehículos no necesitan ser simples ni estáticos para jugar un papel representacional.

La observación de van Gelder es importante. No considera que haya mostrado que no hay representaciones internas: sólo que *pudiera* no haber ninguna, y que, si las *hay*, pueden tomar una forma muy diferente de la que esperábamos una vez. He intentado mostrar que algunas de las consideraciones escépticas más específicas que él avanza (sobre la potencial complejidad de las interacciones agente-entorno) fallan (y deben fallar) para conectar con el argumento pro-representacionista original: un argumento fundado directamente en nuestras capacidades para la razón *ambientalmente desacoplada*. La opción representacionista reví-

49 Knierim y Van Essen, "Visual Cortex", 150-5.

50 Van Gelder, "What Might Cognition Be?".

sionista, no obstante, permanece abierta, atractiva y cada vez más evidente en las reales aplicaciones científicas cognitivas<sup>51</sup>.

En resumen, nuestra visión de la razón biológica está cambiando rápidamente. Hay un énfasis creciente en las economías computacionales permitidas por la acción en el mundo real y un creciente aprecio por el modo como estructuras mayores (de agentes y artefactos) sirven de andamio y transforman el espacio de la razón individual. Esas fuerzas gemelas convergen en una explicación bastante más minimalista del procesamiento cognitivo individual –una explicación que tiende a rehuir los ricos modelos polivalentes internos neutros a la acción y las formas sentenciales de la representación interna. Tal minimalismo, no obstante, tiene sus límites. A pesar de algunos argumentos algo pretenciosos, no hay actualmente ninguna razón para poner en duda la visión rectora de agentes individuales como localizaciones de modestas representaciones internas y usuarios de una variedad de modelos internos del mundo. Mejor que *oponer* el representacionismo a la dinámica interactiva, se debería abrazar una visión más amplia del ámbito representacional interno y buscar las continuidades cruciales entre las estrategias conductuales fuertemente acopladas y el espacio más «cartesiano» de la razón desacoplada ambientalmente. Nuestro premio será una mejor concepción de la propia acción racional.

51 Ver los trabajos en Port y Van Gelder (eds.), *Mind as Motion*.