

Vida artificial y filosofía

Manuel de Pinedo García

The Philosophy of Artificial Life, de MARGARET A. BODEN (ED.). OXFORD, OXFORD UNIVERSITY PRESS, 1996, viii + 408 pp., £11.99

El nuevo volumen de la colección *Oxford Readings of Philosophy* está dedicado al área más reciente de la filosofía, y quizá también la más interdisciplinaria. El estudio de la vida artificial¹ sólo fue bautizado como tal hace diez años. Ha habido tiempo desde entonces para que surjan una gran cantidad de congresos, de revistas especializadas, e incluso de departamentos universitarios dedicados a su investigación.

Mi recensión de este libro será inusual. Tengo la impresión de que el área de investigación del que se ocupa este libro y la reflexión filosófica que la acompaña y en gran medida le da forma, no son muy conocidas en España. Por otra parte, llevo dos años trabajando con la editora y con algunos de los colaboradores de esta colección, y no puedo evitar sentir un entusiasmado compromiso con su contenido. Quizá muchos filósofos se resignen a mirar con nostalgia a nuestro glorioso pasado, cuando la reina de todas las ciencias se llamaba filosofía natural. Sin embargo, cada vez que surge una disciplina nueva, la inspiración y la guía es filosófica. *A-Life* es un nuevo ejemplo y, seguramente, uno particularmente interesante.

El uso de la palabra “vida” puede resultar sorprendente junto al adjetivo “artificial”. Sin embargo, no debería ser así si queremos evitar una definición esencialista de lo que es la vida. Ni siquiera la misma biología puede ofrecerla, ya que en ella las fronteras también son vagas, como es el caso de los virus. Como señala Margaret Boden en la introducción, el concepto de vida cubre una serie de propiedades de las cuales algunas son filosóficamente problemáticas: autoorganización, emergencia, autonomía, crecimiento, desarrollo, adaptación, reproducción, evolución, etc. Sólo conocemos una forma de vida, la terrestre, pero no hay nada que impida considerar vida cualquier cosa que tenga estas propiedades.

Algunos filósofos han observado con alarma que el programa de investigación de la inteligencia artificial se parece peligrosamente al paradigma filosófico que va desde Descartes hasta Husserl y el positivismo lógico, pasando por Kant. La bancarrota de este paradigma, que intenta explicar la naturaleza del ser humano y su relación con el mundo en términos intelectualis-

tas, y su superación en las filosofías de Heidegger, el segundo Wittgenstein, Ortega y Gasset o Merleau-Ponty, no ha tenido el eco deseado en el trabajo de simulación y producción de inteligencia artificial. Así, por ejemplo, Hubert Dreyfus emprendió a principios de los setenta una cruzada contra el resurgimiento del intelectualismo tanto en la IA como en la filosofía de la mente de corte funcionalista y cognitivista. En su artículo “Making the Mind versus Modelling the Brain: Artificial Intelligence Back at a Branch Point” [DREYFUS, H. L. y DREYFUS, S. E. (1992), “Making the Mind versus Modelling the Brain: Artificial Intelligence Back at a Branch Point”, en Boden, M. A. (ed.), *The Philosophy of Artificial Intelligence*, Oxford, Oxford University Press, pp. 309-33] y en el libro editado por él, *Husserl, Intentionality, and Cognitive Science* [DREYFUS, H. L. y HALL, H. (eds.) (1982), *Husserl, Intentionality, and Cognitive Science*, Cambridge, MA, The MIT Press], Dreyfus mantiene que la tradición racionalista se ha visto puesta empíricamente a prueba y ha fracasado. Toda la filosofía occidental se mueve alrededor del procesamiento de información simbólica. Para Heidegger, la filosofía tradicional se centra en “hechos sobre el mundo, mientras que ‘deja de lado’ al mundo como tal” [p. 318]; esto es, la filosofía ha ignorado o distorsionado los contextos cotidianos de la actividad humana. (Para Heidegger, el filósofo que más se acercó a la comprensión de este hecho fue Aristóteles.) Así, por ejemplo, Platón entendió las matemáticas e incluso la ética como una aplicación de las reglas descontextualizadas de una teoría aprendida en otra vida, ajena al mundo cotidiano. Leibniz generalizó esta concepción, extendiéndola a todas las formas de actividad inteligente.

El segundo Wittgenstein y el primer Heidegger cuestionaron esta tradición y propusieron una concepción más holista de la inteligencia que reconociera la importancia de las prácticas cotidianas. Heidegger ya señaló a Husserl que existen otras formas de “encontrarse con” las cosas además de relacionarse con ellas como objetos definidos por un conjunto de predicados. La actividad humana es una combinación de habilidades y prácticas que no son estados intencionales y así, *a fortiori*, no tienen ningún contenido representacional que haya que ser explicado en términos de elementos y reglas. El mismo Husserl llegó a aceptar esto al final de su vida, cuando reconoció que la fenomenología es una tarea infinita.

Este mismo problema fue reflejado en los primeros estadios de la investigación en IA. Por ejemplo, se pensó que el “robot” SHRDLU de Winograd era un primer paso en el proyecto de combinar “micro-mundos” para alcanzar una comprensión global del mundo real. Pero, como diría Heidegger, un universo (un conjunto de hechos interrelacionados) no constituye un mundo. Un universo permanece carente de significado, mientras que el mundo adquiere significado en el contexto de nuestra práctica y el ámbito cultural donde se realiza. El problema del conocimiento de sentido común (“*frame problem*”)² ha bloqueado cualquier progreso en el campo de la inteligencia artificial teórica.

La mayor parte de la investigación en *A-Life* no está sujeta a la misma crítica ni a los mismos problemas. La *IA* tradicional, y los estudios filosóficos y psicológicos que la justifican y se apoyan en ella (cognitivismo, funcionalismo)³, tienen casi siempre un compromiso con representaciones que median entre el agente y el mundo. Éstas son planteadas como hipótesis para explicar nuestra vida mental, normalmente ignorando el papel que el cuerpo y el espacio corporal juegan en la configuración de ésta⁴. En gran parte, el trabajo en *A-life* no comparte este compromiso.

El programador tradicional es un ser omnipotente que estipula fines y que crea modelos formales abstractos del comportamiento. La máquina debe seguir estos modelos utilizando reglas preconcebidas por él. Ahora se ve sustituido por alguien que juega un papel mucho más parecido al de la madre naturaleza. De entre un gran número de programas producidos al azar, el programador selecciona aquellos que más se acercan a realizar lo que se espera de ellos. Por ejemplo, en una simulación de un robot (o “mobot”, o “animat”, como se les ha empezado a llamar) que debe evitar los obstáculos que se encuentra para atravesar una habitación buscando un punto de luz, aquellos programas que lo lleven más lejos son seleccionados y se reproducen cruzándose con otros de los programas seleccionados e incorporando mutaciones. Generación tras generación el programador lleva a cabo la misma operación hasta que casi todos o todos los programas son eficaces. Incluso para las tareas más simples es muy difícil o imposible determinar cómo logra el programa el éxito⁵. Sin embargo, con estos métodos (que usan “algoritmos genéticos” y constituyen la “robótica evolucionista”) se consiguen realizar las tareas más sencillas que se le habían resistido a la *IA* tradicional y se confirma algo que ya se sabía desde hace tiempo: que de nuestras acciones las que más nos pasan desapercibidas (subir una escalera, abrocharse un botón, evitar una mesa, andar) son las más difíciles de formalizar.

El estudio de la inteligencia que tiene lugar en la *A-life* es de inspiración biológica. Como señala Rodney A. Brooks, uno de los pioneros de la nueva robótica, la inteligencia humana es demasiado compleja para descomponerla de forma correcta. Sólo podremos pensar en ello cuando hayamos estudiado mejor niveles de inteligencia más simples. De acuerdo con él, cuando hacemos esto vemos que las representaciones y los modelos del mundo están de más, y que es mejor usar el mundo como su propio modelo⁶. El movimiento, la visión aguda y otras habilidades relacionadas con la supervivencia proporcionan la base para el desarrollo de la inteligencia, y sin embargo el estudio de estas puede hacerse sin necesidad de postular representaciones. En lugar de dividir el organismo o el robot en un sistema central y una serie de sistemas periféricos que dependen de éste (aquél percibe y actúa por medio de éstos), se divide en subsistemas productores de actividad, relacionados entre sí de forma compleja.

De las ideas de Brooks se desprende un modelo de acción distinto del tradicional. La conducta se puede dividir en una serie de habilidades y actividades orientadas hacia tareas particulares: caminar, sentarse, evitar un obstáculo de una determinada forma, encender el equipo de música, etc. Cada una de ellas tiene sus propios requisitos sensoriales y de control, y actúan en paralelo. El mundo exterior funciona, como hemos visto, como memoria, y éste, junto con los mecanismos de control que coordinan las diferentes habilidades, determina cuál o cuáles de ellas se ejecuta. La representación se sustituye por control. Al igual que el conductismo, este modelo niega la existencia de un fantasma de la máquina. Pero a diferencia de éste no tiene miedo de postular estructuras internas⁷.

La inteligencia humana, incluso en sus formas más abstractas, es un mecanismo de supervivencia producto de la selección natural⁸. La relevancia para su estudio de otros mecanismos y del medio donde esta selección tiene lugar es enorme. Es interesante la recuperación del término "*Merkwelt*", debido al biólogo de principio de siglo Jakob von Uexküll. Es usado para referirse al hecho de que el mundo perceptual de una criatura es relativo a la especie, es decir, a sus posibilidades perceptivas, su morfología y sus modelos de acción en el mundo. Esta idea ya influyó en su momento en la noción orteguiana de circunstancia y recientemente en la teoría de Dawkins del fenotipo extendido.

Las consecuencias de este enfoque radicalmente naturalista para la filosofía de la mente son, a mi entender, enormes. El funcionalismo imperante es la recapitulación teórica de un proyecto de investigación que tiene en su núcleo una concepción intelectualista de la mente. Las posibilidades de salir de lo fenoménico y reconstruir el mundo a partir de ello parecen cortadas por argumentos humeanos. Si se toma el mundo como punto de partida, no parece que haya que seguir temiendo a estos argumentos.

El libro se compone de una introducción y de quince artículos, algunos de ellos escritos especialmente para este volumen. La introducción presenta una amplia panorámica de la disciplina al tiempo que contextualiza las quince contribuciones, siguiendo el mejor estilo de esta colección. El primer capítulo es el ya clásico "Artificial Life" de Chris Langton, que fue en gran medida el acta fundacional del programa de investigación al tiempo que su bautismo como tal. Es el complemento perfecto para la introducción. Esta última pone el acento en las cuestiones filosóficas que han surgido en los últimos años alrededor de este campo, en su relevancia para la biología y para la sociedad en general. El artículo "Artificial Life" es una magnífica historia de los orígenes de la disciplina, de su relación práctica con la IA y con la biología teórica y matemática⁹, y está llena de esclarecedoras explicaciones de los conceptos técnicos más duros. Langton defiende una tesis compartida por muchos de los colaboradores de este libro: la vida es una propiedad emergente, una propiedad de la "forma" y no de la materia, es decir un resultado de la organización

de la materia más que algo inherente a la materia misma¹⁰. La palabra “artificial” se refiere a los componentes de la instanciación, no a los procesos emergentes. Cualquier organización que tenga el mismo papel funcional que las biomoléculas de los sistemas vivos naturales manifestará vida genuina.

Tras estos textos introductorios, y un artículo de Margaret Boden (“Autonomy and Artificiality”) del que hablaré más adelante, la segunda parte contiene ejemplos de la investigación en *A-life*. Alguno de ellos ya aparecen brevemente explicados en el trabajo de Langton, incluido el más interesante a mi gusto, el *Tierra Simulator* de Thomas Ray. El mismo Ray, en el capítulo 3, “An Approach to the Synthesis of Life”, ofrece una apasionante descripción de este experimento junto con algunas reflexiones sobre sus consecuencias. Para el autor es accidental que hasta ahora el estudio de la vida haya tenido que limitarse a las formas basadas en el carbono. *A-life* puede compensar esta limitación en su intento de sintetizar, que no simular, vida. El simulador Tierra contiene primeramente un solo programa con 80 instrucciones que puede autoreproducirse. Este programa examina el lugar que ocupa en la memoria del ordenador virtual que le sirve de entorno¹¹, y reserva un espacio igual de la memoria para instalar un programa hijo, idéntico a sí mismo, una y otra vez. Los hijos tienen un comportamiento similar, hasta que el 80% de la memoria está ocupada y una “segadora” (*reaper*) entra en acción y elimina los programas más antiguos. Una baja tasa de mutación (errores de copia) es permitida. Pronto la diversidad de genotipos y de tamaños del genoma (del número de instrucciones del programa) estalla. Los programas evolucionan y con ellos la diversidad ecológica. Surgen parásitos que no pueden reproducirse sin utilizar otros programas, hiper-parásitos que utilizan a los parásitos para reproducirse a sí mismos, programas que sólo pueden reproducirse socialmente y que a su vez se ven invadidos por hiper-hiper-parásitos. Después de varios días de simulación pueden aparecer cientos de grupos de tamaño y hasta 30.000 genomas distintos, algunos de enorme tamaño (más de 20.000 instrucciones). Algunos tamaños son mucho más evolutivos que otros, así como algunas combinaciones de organismos/programas. Según el autor, esta simulación sugiere que la interacción con otras especies que evolucionan proporciona la principal fuerza a la evolución. La discusión de Ray sobre micro y macro-evolución, emergentismo, diversidad y mutación es apasionante y de gran interés para la biología. De todas formas, como hemos visto, el autor piensa que el experimento es biología, y no meramente un instrumento de uso biológico.

La tercera parte del libro, junto con el segundo capítulo, es donde la discusión filosófica de mayor profundidad tiene lugar, ya que en ellos se analizan las estrategias explicativas de la *A-life* y su relevancia para la filosofía y las ciencias cognitivas. El segundo capítulo (“Autonomy and Artificiality”) llama la atención sobre la insistencia que los investigadores de *A-life* ponen en el concepto de autonomía como señal de identidad con respecto a la *IA*.

Según la autora, que mantiene el compromiso con esta última, las manifestaciones más complejas de la autonomía (auto-conciencia y razonamiento) siguen siendo mejor explicadas por la IA tradicional. Esto no puede ser una sorpresa. *A-life* persigue un modelo explicativo de abajo a arriba, y es un proyecto recién nacido. La “nueva IA” (*nouvelle AI*), y en especial la “robótica situada” (*situated robotics*), está cada vez más interesada en el diseño de agentes autónomos mucho más simples que la mente humana.

El capítulo 7, “From Robots to Rothko: The Bringing Forth of Worlds” de Michael Wheeler, critica el cartesianismo de la ciencia cognitiva ortodoxa desde una perspectiva muy próxima a la hermenéutica¹². Su argumentación en favor de la compatibilidad entre el naturalismo y la hermenéutica es, a mi entender, muy convincente y en cierto sentido acerca al autor a las ideas de Donald Davidson sobre interpretación radical y monismo. La separación entre acción y percepción se borra cuando se trata de agentes autónomos en el contexto de un proceso de conducta adaptativa¹³. El cognitivismo tradicional, restringido según Wheeler a la dicotomía sujeto-objeto, al representacionalismo y al computacionalismo, no está en condiciones de explicar la emergencia de procesos complejos ni las igualmente complejas relaciones entre organismos y medio ambiente.

Sin embargo, a las filas del representacionalismo no le faltan argumentos y éstos normalmente proceden del sentido común filosófico, encabezado por Jerry Fodor: los filósofos nos pasamos la vida pensando y hablando, y llegamos a creer que éste es el rasgo más característico del ser humano. En verdad no lo es menos la tendencia que nuestros cuerpos tienen a extenderse más allá de sí mismos con el uso de instrumentos fabricados por ellos. David Kirsh, en el capítulo 8, “Today the Earwig, Tomorrow Man?”, expone sus dudas acerca de la posibilidad de explicar las conductas de alto nivel sin un estudio de las representaciones. Comprender el papel del pensamiento en la acción obliga a hablar de abstracciones, de conceptos. Cuando una tarea exija conocimiento sobre el mundo que sólo se obtiene por razonamiento o memoria y no por percepción, no puede considerarse determinada por el entorno. Por ejemplo cuando otros agentes están involucrados y es necesario predecir su conducta. Hay muchas tareas que tienen necesidad de contrafácticos, y estos no son “percibibles”.

Una respuesta moderada a esta objeción puede encontrarse en el siguiente capítulo “Happy Couplings: Emergence and Explanatory Interlock”, de Andy Clark, y otra más radical en el posterior a éste, “In Praise of Interactive Emergence, Or Why Explanations Don’t Have to Wait for Implementations” de Horst Hendriks-Jansen. Clark mantiene que existen dos proyectos explicativos en el estudio del conocimiento. El nuevo paradigma que surge del estudio de los sistemas agente/entorno, es decir, del estudio de agentes encarnados e incrustados (*embodied and embedded*) no desplaza totalmente los modelos explicativos tradicionales que utilizan representaciones y compu-

tación. El interés de Clark no está tanto en eliminar la explicación representacional como en situarla en contexto. Las representaciones, y en particular los conceptos, han de naturalizarse y para ello es preciso partir de representaciones no-conceptuales, y en última instancia de procesos no representacionales. Los mecanismos de formación preferidos por Clark son redes conexionistas. No nos engañemos, el contenido no-conceptual no deja de ser representacional. Sin embargo, el intento de hacer compatible un interpretacionismo cercano a Davidson con la noción de contenido no-conceptual es hoy uno de los más prometedores de la filosofía de la mente. Hendriks-Jansen reconoce que los humanos usan conceptos, pero cree que no debe partirse de ellos para explicar nuestros procesos cognitivos en general, como ha demostrado la falta de progreso en *IA* y en los análisis de arriba a abajo. Los nuevos descubrimientos sugieren que el significado emerge a partir de interacciones dinámicas que no son significativas.

Los tres capítulos siguientes contienen discusiones filosóficas más tradicionales del concepto de vida, y de las enseñanzas de Aristóteles, Dewey y Spencer. Y los dos últimos relacionan *A-life* con el funcionalismo en filosofía de la mente y de la biología. No dejan de ser interesantes, pero la ya excesiva longitud de mi comentario no me permite referirme a ellos. Espero haber colaborado en estas páginas a despertar el interés por este área que, como he dicho al principio, tan importante me parece para la filosofía actual.

School of Cognitive and Computing Sciences
University of Sussex
Falmer, Brighton, East Sussex, BN1 9QN, UK.

NOTAS

¹ La abreviatura que se ha impuesto en la literatura es *A-Life*. Para la inteligencia artificial, por el contrario, usaré la abreviatura española, *IA*. Uso la abreviatura inglesa de vida artificial, pues aún no se ha generalizado un equivalente castellano.

² Mientras que desde el principio no ha habido gran dificultad en producir programas que ejecuten tareas formales incluso mejor que los propios humanos, las actividades aparentemente más simples (por ejemplo el simulacro de un pie que sube un escalón, o la respuesta a preguntas sobre situaciones sencillas, que sin embargo involucran tener en cuenta qué es relevante para la pregunta en la situación) se han resistido a los intentos de solución por parte de los programadores.

³ Que, en su conjunto, han llegado a ser llamados *GOF AI* (“Good old-fashioned Artificial Intelligence”, “la buena inteligencia artificial al viejo estilo”), usando el simpático acrónimo de John Haugeland [HAUGELAND, J. (1985), *Artificial Intelligence: The Very Idea*, Cambridge, MA, The MIT Press].

⁴ Las ideas de Merleau-Ponty aparecen en la filosofía de la vida artificial casi tanto como las de Heidegger. La dura crítica al intelectualismo con la que comienza

su *Fenomenología de la percepción*, su teoría sobre la corporeidad, y sus análisis del ciclo acción-percepción han sido estudiados por varios autores. Sin embargo, el impacto de Heidegger es impresionante. Que el mismo Winograd enseñe a Heidegger en sus cursos de ciencias computacionales en Stanford ha dejado de ser una anécdota aislada.

⁵ Por supuesto que esta explicación es simplista y, en parte, inexacta. No obstante, me parece que puede dar una idea aproximada.

⁶ BROOKS, R. A. (1991), "Intelligence without Representation", *Artificial Intelligence*, vol. 47, pp. 139-59.

⁷ En el capítulo 8, de David Kirsh, este modelo es criticado, como veremos más adelante.

⁸ John Maynard Smith, uno de los principales artífices de la "nueva síntesis" del evolucionismo y la genética, llama la atención sobre la lección que hemos de aprender de los estudios de la biología, y a ello dedica su contribución a este volumen ("Evolution — Natural and Artificial").

⁹ Con respecto a la primera, se discute el adaptacionismo y la tesis de Gould y Lewontin del equilibrio interrumpido (*punctuated equilibrium*). Sobre este debate, y sobre las conclusiones desmesuradas y esencialistas a las que ha conducido, ver DENNETT, D. C. (1987), "Intentional Systems in Cognitive Ethology", en *The Intentional Stance*, Cambridge, MA, The MIT Press, y DENNETT, D. C. (1995), *Darwin's Dangerous Idea*, Londres, Allen Lane, donde se tratan muchas de las cuestiones que han surgido en el campo de la *A-Life*. El ejemplo del dilema del prisionero y la explicación de la teoría de los sistemas dinámicos sirve de ilustración de los problemas de la biología matemática.

¹⁰ Una tesis similar sirve a Elliot Sober en el capítulo 14, "Learning from Functionalism — Prospects for Strong Artificial Life", para mantener que el funcionalismo es tan defendible en la filosofía de la biología como en la filosofía de la mente, al permitir la implementación múltiple.

¹¹ El uso de un ordenador virtual en lugar de uno real está destinado a que este programa y sus descendientes no infecten el ordenador y se salgan de madre.

¹² Una hermenéutica opuesta a la de Gadamer y a su idea de la "lingüística universal". El autor, un gran conocedor de Heidegger que es capaz de hacer comprender las ideas del filósofo alemán a audiencias sin la menor formación filosófica (toda una hazaña), señala que el concepto heideggeriano "*Rede*" no debería ser traducido por "habla" (*talk*), ya que la *Rede* no tiene que ser lingüística.

¹³ El famoso trabajo de Gibson sobre la percepción está basado en consideraciones muy parecidas. Ver GIBSON, J. J. (1979), *The Ecological Approach to Visual Perception*, Boston, Houghton Mifflin.