

LA ESCALERA DE LA COMPLEJIDAD VIDA ARTIFICIAL II

● ÁNGEL DE LA ENCARNACIÓN GARCÍA BAÑOS ●



Universidad
del Valle

Programa  Editorial

Este libro es una mezcla de ensayo e investigación acerca de cómo usar las herramientas presentadas en el libro anterior con el objetivo de crear sistemas de complejidad creciente que nos llevan hacia entes con libertad, inteligencia y consciencia. En particular, hago un desarrollo completamente nuevo sobre este último tema, la consciencia, explicada desde un punto de vista computacional. Ya tengo un artículo publicado sobre ello, otro sometido a revisión en una revista y otro en curso de escribirlo. Quiero resaltar la importancia del tema, por lo que es preciso su divulgación. La consciencia ha sido el último rincón oscuro de la ciencia durante demasiados años, y no ha habido propuestas interesantes para resolverlo hasta ahora.

Se explica la complejidad, como se puede definir y medir. Introduce los conceptos de emergencia, cambio de fase, auto organización, si la complejidad puede crecer indefinidamente y si es un fenómeno gradual o no. Como propuesta original, presento un algoritmo evolutivo que sirve para generar complejidad de manera universal.

Lo siguiente que aparece en la escalera de la complejidad es la inteligencia, para lo cual no se requiere mucho. Ofrezco una definición de inteligencia basada únicamente en la capacidad de predecir. Se es más inteligente cuanto mejores predicciones se hacen respecto al sucesos futuros y cuando más profundamente en el tiempo se pueda hacer. La inteligencia es entonces un complementario de la libertad (no contrario, sino complementario, pues se ayudan mutuamente a crecer). Sin olvidar los recientes avances en Deep Learning de la Inteligencia Artificial, se presenta una propuesta de los distintos algoritmos que usa la inteligencia. Además se plantea que toda inteligencia es colectiva.



LA ESCALERA DE
LA COMPLEJIDAD
VIDA ARTIFICIAL II

ÁNGEL DE LA ENCARNACIÓN GARCÍA BAÑOS



Colección Ciencias Naturales y Exactas

García Baños, Ángel de la Encarnación

La escalera de la complejidad. Vida artificial II / Ángel de la Encarnación García Baños. -- Cali : Programa Editorial Universidad del Valle, 2019.

356 páginas : ilustraciones ; 24 cm. -- (Colección ciencias naturales y exactas)

Incluye bibliografías.

1. Realidad virtual 2. Inteligencia artificial 3. Simulación por computadores 4. Complejidad computacional I. Tít. II. Serie. 006.3 cd 22 ed.

A1644720

CEP-Banco de la República-Biblioteca Luis Ángel Arango

Universidad del Valle

Programa Editorial

Título: *La escalera de la Complejidad. Vida artificial II*

Autor: Ángel de la Encarnación García Baños

ISBN PDF: 978-958-765-986-3

DOI: 10.25100/peu.151.73

Colección: Ciencias Naturales y Exactas

Primera edición

Rector de la Universidad del Valle: Édgar Varela Barrios

Vicerrector de Investigaciones: Jaime R. Cantera Kintz

Director del Programa Editorial: Omar Díaz Saldaña

© Universidad del Valle

© Ángel de la Encarnación García Baños

Diseño y diagramación: Alejandro Soto Perez

Corrección de estilo: María Camila Cuenca O.

Este libro, o parte de él, no puede ser reproducido por ningún medio sin autorización escrita de la Universidad del Valle.

El contenido de esta obra corresponde al derecho de expresión del autor y no compromete el pensamiento institucional de la Universidad del Valle, ni genera responsabilidad frente a terceros. El autor es el responsable del respeto a los derechos de autor y del material contenido en la publicación, razón por la cual la Universidad no puede asumir ninguna responsabilidad en caso de omisiones o errores.

Cali, Colombia, septiembre de 2019

LA ESCALERA DE
LA COMPLEJIDAD
VIDA ARTIFICIAL II



Colección Ciencias Naturales y Exactas

Ángel de la Encarnación García B.

Madrid, España en 1960. Completó sus estudios de Ingeniero de Telecomunicación en 1985 en la Universidad Politécnica de Madrid. En 1987 se afincó en Colombia trabajando independientemente en su propia empresa. En 1993 ingresa a la Universidad del Valle, como profesor de Electrónica en temas que van desde los microprocesadores, interfaces, lógica digital, FPGAs, VHDL y tiempo real. En esa época realiza su doctorado en la Universidad Politécnica de Valencia, España, graduándose en 1999. A partir de allí se traslada al departamento de Ingeniería de Sistemas y Computación donde trabaja en temas de programación orientada a objetos, sistemas operativos, vida artificial y computación evolutiva. Fundó el laboratorio EVALAB (Evolución y Vida Artificiales) y el grupo de investigación GUIA en Inteligencia Artificial. Hizo parte durante varios años del grupo de filosofía de la mente MENTIS. Autor de múltiples artículos publicados en revistas y actas de congresos.

CONTENIDO

PREFACIO	11
INTRODUCCIÓN	19
¿Para qué sirve la vida artificial?	19
Muchos mundos	20
Mundos reales y mundos virtuales	23
Resumen	26
Referencias	27

CAPÍTULO 1

COMPLEJIDAD	29
¿Por qué se genera complejidad de forma espontánea?	35
Complejidad versus entropía	41
¿Cómo medir la complejidad de un sistema?	46
La gradualidad y la emergencia	54
¿La complejidad puede crecer indefinidamente?	57
Emergencia, transiciones de fase y autoorganización	63
Generador de complejidad evolutivo	81
Resumen	87
Referencias	93

CAPÍTULO 2

LIBERTAD	97
¿Cómo conseguir libertad?	106
Libertad y emergencia	108
Libertad y transiciones de fase	108
Resumen	109
Para saber más	109
Referencias	111

CAPÍTULO 3

COMPUTACIÓN COMPLETA	113
Grafos	114
Computación universal	118
¿Qué es un simulador?	119
Límites computacionales	127
¿Qué importancia tienen estos límites?	151
Resumen	160
Para saber más	161
Referencias	161

CAPÍTULO 4

VIDA	163
Vida es inteligencia	167
Vida es reproducción	169
Vida es evolución	174
Vida es competencia, pero también cooperación	189
Resumen	201
Para saber más	202
Referencias	203

CAPÍTULO 5

INTELIGENCIA	207
Definición de inteligencia	213
Medición de la inteligencia	218
¿Somos los humanos inteligentes?	221
Clasificación de los problemas de IA	224
¿Cómo razonan los robots?	234
¿Cómo razonamos los humanos?	238
Inteligencia colectiva	247
Nueva perspectiva	258
Resumen	260
Para saber más	262
Referencias	264

CAPÍTULO 6

CONSCIENCIA	269
Una propuesta computacional para entender la consciencia	281
El “hard problem”	292
La consciencia es la burocracia del cerebro	302

Resumen	306
Para saber más	307
Referencias	308

CAPÍTULO 7

CONCLUSIONES	313
Libertad e inteligencia	320
Otra forma de modelar la reproducción	321
Despedida	324
Resumen	327
Referencias	328

CAPÍTULO 8

SOLUCIÓN A PROBLEMAS DE INGENIO	331
Las 12 monedas	331
¿Cuál es el sistema más complejo?	333
Algoritmo descriptor	333
El bar “El Farol”	334
Teorema de la amistad	334
Galletas para todos	335
Acertijo MU	336
Dos autoreferencias divertidas	337
El barbero de Sevilla	337
Paradoja de Richard	338
Lo que Henry decidió	339
Campeón en reproducción	341
Comunicaciones en 2D	342
En la consulta médica	342
¿De cuántas formas sabes multiplicar 15x18?	343
Salvar el foso	344
NIM	344
Detectar engaños	345
Las tres puertas	345
Aplanando la circunferencia	347
Punto fijo de Palmira a Tuluá	347
Arrugando papeles	348
La vela	349
Para saber más	350
Referencias	351
GLOSARIO	353

PREFACIO

La vida artificial es una disciplina relativamente nueva que surgió en 1987 en el primer congreso sobre el tema. Fue definida por su organizador Christopher Langton como “la vida, no como es, sino como podría ser”. A pesar de que llevamos trabajando durante tres décadas, considero que todavía es una disciplina poco madura pues le falta alcanzar su objetivo principal: crear vida artificial en el computador. Esto lo menciono sin pretender menoscabar los grandes avances que se han dado en el campo y que veremos a lo largo de este libro, pero lo cierto es que todavía no hemos logrado un gran éxito, como desarrollar un robot autónomo que se desenvuelva por su cuenta y que podamos decir que “es como si estuviese vivo”.

Pero ya casi.

El interés por esta disciplina ha ido en aumento y desde su comienzo se han escrito algunos libros muy buenos, aunque también podríamos decir que no han sido muchos. Cuando uno los lee se queda deslumbrado ante una variedad de técnicas, conceptos y ejemplos sorprendentes, pero que no se sabe muy bien a qué apuntan o por qué están allí. ¿Qué tiene que ver el caos con los autómatas celulares? ¿Usando ambos se logrará vida artificial? ¿Por qué? Y, por cierto, ¿qué es la vida artificial y cuáles son sus aplicaciones e importancia, aparte de hacer dibujos animados muy realistas? Estas y muchas más son preguntas que surgen en los interesados en la vida artificial.

Y ahora, en lo que atañe más directamente al texto que propongo, ¿en qué se diferencia de otros que tocan el mismo tema? El tema es tan amplio que se hizo necesario separarlo en dos libros (aunque podrían ser más en el futuro): en el primero, titulado *Herramientas para construir mundos*, contemplo todas

las técnicas en detalle para que sirvan como ayuda a cualquier programador, informático o científico de la computación al desarrollar su propio *software*. A continuación presentaré una visión general de estos tópicos.

En primer lugar se habla de la realimentación, que ocurre cuando las acciones que realizas sobre un objeto de alguna manera se devuelven sobre ti, con lo cual la idea de causa y efecto se desvanece, a la vez que se hace más difícil predecir lo que cada acción va a desencadenar.

En el siguiente capítulo se presentan los fractales, unos objetos muy bonitos, con muchos detalles y colores, que se fabrican a partir de una fórmula matemática tan simple que nadie lo creería. Además, no parecen ser objetos comunes de una, dos o tres dimensiones —que corresponderían aproximadamente a una cuerda, una hoja de papel o una mesa—, sino que pueden tener otras dimensiones que no son números enteros, como 0.8 o 2.45.

Luego hacen su aparición los fenómenos caóticos, para los que es imposible predecir a largo plazo cuál va a ser su comportamiento, a pesar de conocerse todos sus detalles de funcionamiento y no contener ningún proceso que dependa del azar. El ejemplo más conocido es el clima y, gracias a ello, tenemos tema de conversación sobre cómo se han equivocado en las predicciones de ayer de la televisión. Sin embargo, existen muchos otros fenómenos caóticos que van desde las cotizaciones de acciones en la bolsa hasta nuestros propios movimientos al caminar, e incluso fenómenos fisiológicos como los latidos del corazón.

Pasamos luego a las leyes de potencia, las cuales son distribuciones de probabilidad que ocurren tanto en el ámbito humano como en las ciencias naturales. Se caracterizan porque hay unos pocos fenómenos dominantes que consumen la mayor parte de los recursos de un sistema. Muchas veces se enuncia de esta forma: el 20% de los clientes de un negocio generan el 80% de los beneficios; el 20% de las líneas de código de un programa son las causantes del 80% de los *bugs*; el 20% de las páginas web son visitadas por el 80% de los navegantes; el 20% de los terremotos producen el 80% de las víctimas. Son innumerables.

Estos cuatro conceptos, realimentación, fractales, caos y leyes de potencia, están entrelazados. Siempre que se da uno aparecen los demás —aun cuando la realimentación es el más fundamental—, y todos generan mucha complejidad a partir de muy poco. Son procesos creativos.

Los algoritmos evolutivos descritos en el sexto capítulo, son una generalización del proceso de la evolución descubierto por Darwin, y se aplican a todo tipo de problemas donde haya que buscar una solución u optimizarla. Y cuando digo “todo tipo” no estoy exagerando, pues sirven para problemas de *software*, de economía, de matemáticas, de diseño industrial, para jugar

y ganar cualquier juego e incluso para crear música o pintura. Posiblemente sean los algoritmos más simples que existen, pero a cambio son muy versátiles y robustos, tanto que se consideran uno de los pilares de la inteligencia artificial. Y aunque hay que admitir su lentitud, eso no es una gran desventaja gracias a la potencia de los computadores actuales. Después de todo, la evolución biológica necesitó varios miles de millones de años para llegar al cerebro humano. Con los computadores actuales se pueden lograr resultados interesantes en cuestión de minutos. Los algoritmos evolutivos son grandes creadores de complejidad, mucho más que los cuatro procesos anteriores. Sin embargo, exigen una condición difícil de lograr de forma espontánea: que los entes que van a evolucionar sean capaces de realizar copias de sí mismos. Todo ello lo analizaremos en detalle.

En el séptimo capítulo se presenta una somera introducción a la teoría de juegos, con el objetivo de entender cómo emerge la cooperación a partir de los procesos evolutivos que, en principio, son todo lo contrario (son competitivos). Veremos que la cooperación genera complejidad en cantidades enormes, mucho más que los algoritmos evolutivos por sí solos.

Por último, veremos los autómatas celulares, entendidos como una gran cantidad de individuos idénticos y muy simples, donde cada uno se comunica solo con sus vecinos. No hay ningún dato global ni tampoco un control central. A pesar de ello, y a partir de comportamientos tan simples, la sociedad en su conjunto puede desarrollar una complejidad todavía mayor a las que hemos enumerado. En particular, estudiaremos que una sociedad así es capaz de sacar copias de sí misma y de hacer cualquier tipo de cómputo.

Todo lo anterior forma parte del bagaje bien conocido en el mundo de la vida artificial, pero quiero presentarlo de una manera nueva que ya estoy mencionando a cada momento: como generadores de complejidad. La importancia de esto radica en la idea intuitiva que tenemos acerca de que los sistemas simples no pueden hacer mucho, mientras que los sistemas complejos pueden moverse, reaccionar a estímulos, adaptarse a su entorno, pensar e incluso ser conscientes de sí mismos (que son temas que veremos en el segundo libro). Por ello necesitamos primero conocer estas técnicas que fabrican sistemas complejos.

También para los programadores presento un ejemplo de algoritmo evolutivo, escrito por mí en lenguaje Ruby y usando metodología BDD (*Behavior Driven Development*) con la herramienta Cucumber. Uso Ruby porque es un lenguaje sencillo y elegante, con mínima sintaxis y gran expresividad. Después de 30 años programando principalmente en C++, que es perfecto pero complicado, y luchando contra muchos otros lenguajes incoherentes, descubrir Ruby ha sido una sorpresa refrescante pues me permite escribir

código conciso con poco esfuerzo. Pero sé que no es un lenguaje muy popular, de modo que en los apéndices también ofrezco una introducción básica a Ruby y Cucumber¹.

En el segundo libro, *La escalera de la complejidad*, exploro los aspectos filosóficos de la vida artificial bajo un hilo conductor que es, como indica el título, la complejidad. Veremos que, conforme aumenta, emergen propiedades nuevas por las que podremos decir que hemos creado un mundo computacional que disfruta de libertad, que está vivo, que es inteligente, o incluso que es consciente. A cada uno de estos temas les dedico un capítulo. Doy un enfoque nuevo a lo que significa libertad e inteligencia. Y presento mi visión de lo que es la consciencia bajo un punto de vista computacional, despojándola de todo misterio. La consciencia tiene varios aspectos que han desconcertado desde siempre a los filósofos, pero todos son entendibles desde esta perspectiva. Quizás sea este el aporte más importante de los dos libros.

Como veremos también, todas estas propiedades son graduales. Hay una cierta inteligencia en una ameba e incluso en un destornillador². Y aunque consideramos que tanto plantas como animales están vivos, en los animales la vida alcanza más posibilidades, como moverse rápidamente. Lo mismo se puede decir de la consciencia que existe en diversos grados en los animales y tal vez incluso en las plantas. De modo que es natural la aparición simultánea de varias propiedades. Si las presento en este orden es porque la consciencia más sofisticada que conocemos requiere un poco más de complejidad que la inteligencia más sofisticada que conocemos, que a su vez requiere más complejidad que la vida, y así sucesivamente. Explicaré que el único salto notable en complejidad se da al pasar de objetos que pueden computar a objetos vivos. Los demás saltos son relativamente pequeños. Y por eso utilizo como hilo conductor la complejidad, que es simplemente una fórmula. Al final, todos estos temas se reducen a matemáticas y, más concretamente, a computación.

1 En el siguiente enlace al depósito de *software* Github se encuentra la información actualizada del libro con las erratas, el *software* y hojas de cálculo desarrolladas para este su realización: <https://goo.gl/EB4gTq>

2 La mera existencia de los destornilladores y sus formas insinúan la existencia de los tornillos. Los tornillos no se podrían manipular sin el destornillador, de modo que el destornillador es como un artefacto de inteligencia congelada, que ayuda a realizar ciertos trabajos. En este sentido, posee una cierta inteligencia aun cuando no esté vivo.

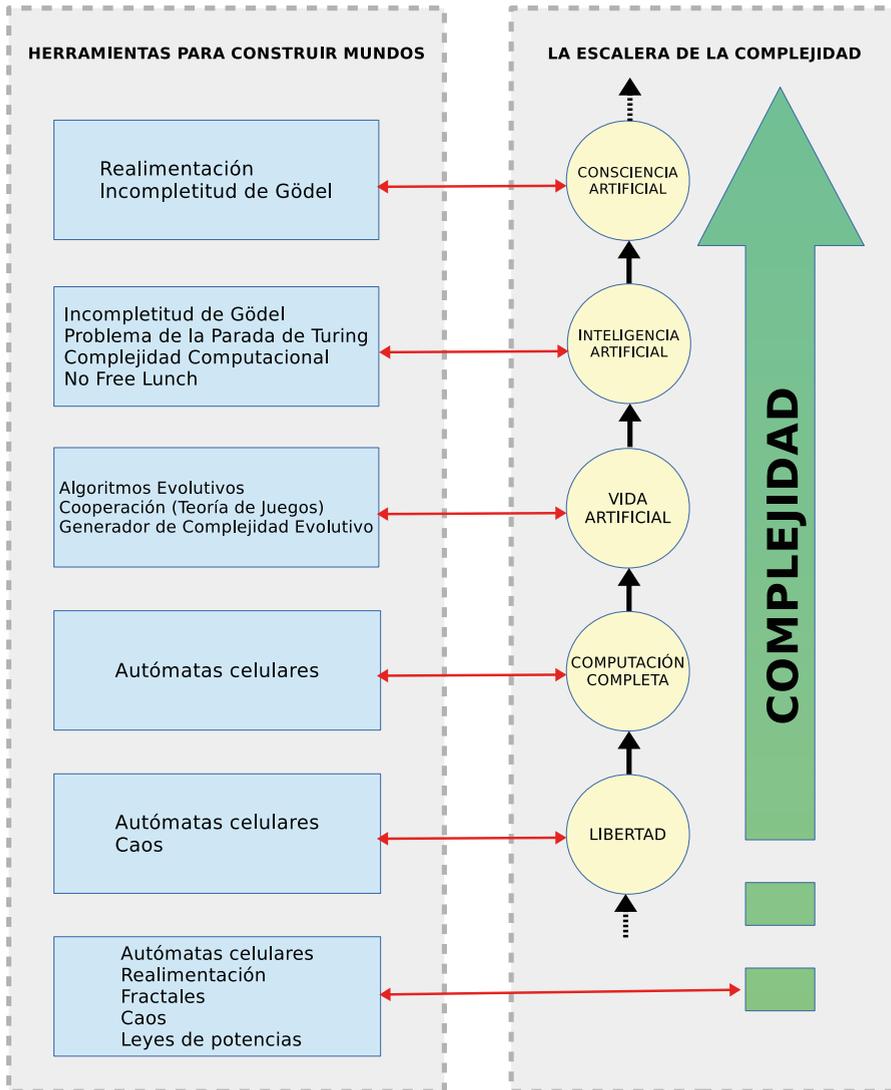


Figura 1. Estructura de los dos libros y relaciones entre ellos

Además, presento una idea nueva: un generador de complejidad evolutivo, que toma varios de los conceptos expuestos y los combina para crear una herramienta potente y general para fabricar sistemas complejos.

En resumen, el primer libro es de sistematización del conocimiento para programadores, mientras que el segundo libro es de investigación, más filosófico. En la figura 1 vemos de abajo hacia arriba los temas que desarrollo en cada libro y la correspondencia entre ellos, es decir, qué herramientas hace falta conocer para entender cómo se caracteriza la complejidad de cada nivel de emergencia.

En ambos libros también incluyo recuadros que contienen información paralela: en color azul están unas pequeñas notas bibliográficas de los personajes más importantes que han contribuido en estos temas; en color verde hay problemas de ingenio que ilustran alguna idea recién presentada o que se va a presentar; y en amarillo están las ideas más importantes que deseo resaltar. Al final de cada libro incluyo las soluciones a esos problemas de ingenio que voy proponiendo para estimular la reflexión sobre ciertos temas.

Verás que con cierta frecuencia menciono cuentos y películas de ciencia ficción. Eso no hace que los libros sean menos rigurosos. Lo que hay que entender es que la investigación tiene dos fases, y cuando formulas una hipótesis, diseñas experimentos y verificas si los resultados están o no de acuerdo con la hipótesis, estás en la segunda fase. No se suele hablar de la primera fase —las ideas iniciales que llevan a plantear la hipótesis—, seguramente porque nadie sabe cómo hacerlo, no hay una metodología que te lleve a las buenas ideas.

Simplemente se requiere inspiración. Y es allí donde las novelas y películas que exploran libremente el futuro sin censuras, sin revisión de pares ni otra atadura, pueden servir como chispa creativa para generar esas buenas ideas. Un momento acertado para capturar las ideas generadas por el inconsciente se da al despertar, cuando aún no se ha logrado vencer del todo el sueño. El lóbulo frontal del cerebro no parece estar activado todavía, censurando los pensamientos para que sobrevivan únicamente los que están acordes con la realidad comúnmente aceptada, en una especie de proceso evolutivo de las ideas. Y, con la censura dormida, eso permite que surjan las ideas más locas. A veces incluso nos sirve estar distraídos en una actividad manual —como lavar los platos— para bajar el umbral de censura. Si quieres continuar investigando en estos temas te animo entonces a que leas mucho y veas cine, especialmente de ciencia ficción dura, y a que laves los platos con mayor frecuencia.

Además, en las referencias incluyo diversos videos, algunos de TED y otros de YouTube. Los primeros son reconocidos por su alta calidad y en ellos participan investigadores de primera línea. Se puede criticar a los segundos, debido a que son publicados libremente sin pasar por una revisión de pares. Y es verdad. Pero lo cierto es que a veces un video o una animación son más esclarecedores que todo el texto que yo pueda escribir aquí, de modo que corro gustoso el riesgo.

El público objetivo son los estudiantes que habitualmente toman mis asignaturas electivas de “Computación evolutiva” y “Vida artificial”, donde el requisito principal es tener experiencia en programación de computadores. No obstante, dada la época en que vivimos donde cualquier persona debería

saber leer, escribir, la aritmética básica y programar, no es descabellado decir que estos dos volúmenes son también de divulgación, asequibles a personas con una mínima formación científica. En cualquier caso, cada libro se puede leer de manera independiente de acuerdo al interés del lector.

Desde el año 2000 trabajo como profesor en la Escuela de Ingeniería de Sistemas y Computación de la Universidad del Valle, dictando varias asignaturas incluidas las ya mencionadas. En estos libros pretendo reunir todo el material de clase, darle una coherencia y mayor profundidad, así como añadir algunos temas nuevos. Para entender rápidamente en qué consisten estas asignaturas te diré que la computación evolutiva abarca muchos algoritmos de optimización inspirados en la evolución de Darwin, aunque aquí los mostraré desde otra perspectiva: algoritmos creativos. Y con la vida artificial se intenta construir mundos artificiales en el computador donde emerjan objetos con comportamientos tan complejos que pueden considerarse vivos, es decir, que se adaptan a su medio y se reproducen, colaborando o compitiendo con otros objetos.

Al pensar en el diseño de mundos artificiales es inevitable la pregunta de si nuestro propio mundo también lo es, o sea, si vivimos dentro de un computador, *a la Matrix*, o en alguna otra variante. Intentaré justificar una respuesta afirmativa a esta pregunta, bastante relacionada con una nueva concepción de la Física llamada *Digital Physics*, de la que haré algunos breves comentarios. El mundo es solo información.

Quiero agradecer a mi esposa Helga Rocío por todo su apoyo durante la escritura de estos libros, por hacer la primera revisión, ayudarme a organizar las referencias y hacer algunos bonitos dibujos. A mi hija Maya por ayudarme con la revisión del inglés de algunos artículos relacionados con este trabajo. Y a mi hermano Rafa por encontrar algunos de los errores más insospechados. El libro lo escribí en España durante mi año sabático del 2016-2017.

Y por ello también agradezco a la Universidad del Valle, pues nos ofrece a los profesores un tiempo para desarrollar ideas sin las presiones académicas y administrativas de todos los días. En la Escuela de Ingeniería de Sistemas y Computación disponemos del laboratorio EVALAB (evolución y vida artificiales), uno de varios donde opera el grupo GUIA de investigación en inteligencia artificial al que pertenezco. En ese laboratorio trabajan o han trabajado varios profesores (Irene Tischer, Víctor Andrés Bucheli Guerrero, Raúl Gutiérrez de Piñérez Reyes, Henry Antonio Saltaren Quiñones, Gabriel Conde Arango) y muchos estudiantes de pregrado, maestría y doctorado cuyos proyectos de fin de carrera se mencionan en el libro. Entre estos estudiantes, ahora ingenieros, quiero agradecer a Cristian Leonardo Ríos

López que me haya cedido muy amablemente varias figuras de su trabajo de grado, y a Eider Falla que me haya permitido mostrar su *software* de la bandera francesa. Las discusiones con todos estos profesores y estudiantes han sido muy enriquecedoras a lo largo de tantos años, y recuerdo con especial agrado las reuniones de los martes en la tarde en el laboratorio y las generales en Piedralinda, debatiendo ideas y compartiendo un tiempo con personas sensacionales. En ese ambiente fue posible escribir el presente libro, por lo que les agradezco mucho a todos ellos su esfuerzo.

Así mismo, durante mi paso por el grupo de filosofía MENTIS de la misma universidad, recuerdo con mucho agrado a profesores y estudiantes de filosofía quienes me iniciaron en el tema de la consciencia, Juan Manuel Cuartas Restrepo, Ómar Díaz Saldaña, Ernesto Enrique Combariza Cruz, Juan Carlos Vélez Rengifo y Germán Guerrero Pino.

También agradezco al profesor Jesús Alexander Aranda Bueno por revisarme el teorema de Gödel. Al profesor Carlos Alberto Mayora Pernía, por revisarme el inglés del primer artículo relacionado con este trabajo. Y al profesor Fabio Germán Guerrero Moreno siempre muy interesado en estos temas y que me hizo una pregunta que me ha mantenido activo buscando su respuesta por varias décadas: ¿cuáles son los límites computacionales fundamentales? A los evaluadores quiero agradecerles por sus generosos comentarios y útiles sugerencias.

Para terminar, doy las gracias a mis amigos José Arturo, Diana Lorena y Víctor Manuel por su apoyo continuado en este y otros proyectos, y por tantas horas dedicadas a conversar sobre ideas interesantes.

Desde luego que si quedan errores deben achacármelos a mí, como siempre se dice en estos casos, pero no por repetirlo deja de ser verdad. Mi principal afán fue presentar una disciplina coherente en las asignaturas que ofrezco y, especialmente, las ganas de aprender más. Espero despertar las mismas ganas a los lectores, aunque seguramente todos nos quedaremos con más preguntas que respuestas, pues es un tema abierto a más investigaciones.

El *software* y las figuras son míos excepto donde se indique lo contrario. Una parte procede de los cursos electivos que dicto habitualmente en la Universidad del Valle y el resto fueron diseñados *ex profeso* para estos libros. Los documentos de texto están escritos con *Libre Office 5.1.6.2* (usando *itálicas* para ecuaciones en el texto, tecnicismos y extranjerismos), corriendo sobre *Ubuntu 16.04LTS* y haciendo *backups* varias veces al día en diferentes medios, usando *Dropbox*, *rdiff-backup* y *git*. Como pueden imaginarse, en mi escuela estamos orgullosos de usar y apoyar el *software* libre.

INTRODUCCIÓN

La vida artificial es el estudio de las propiedades abstractas de los seres vivos, independientemente del sistema físico donde estén implementados. El sustrato de la vida natural es la química del carbono, aunque podría haber otros sustratos químicos, mecánicos, electrónicos o informáticos donde también sea posible crear sistemas que muestren las mismas propiedades. Un ejemplo muy bonito, aunque todavía simple, son los animales de plástico de Theo Jansen que se pasean por las playas. En general, la vida artificial intenta entender de dónde surge la complejidad de estos sistemas, creando herramientas matemáticas y computacionales para poder analizarla, manipularla y sintetizarla.

En el libro anterior hemos visto las herramientas computacionales que nos permiten crear vida artificial. Ahora trataremos de dar una aproximación más filosófica a este tema.

¿PARA QUÉ SIRVE LA VIDA ARTIFICIAL?

Acabamos de enunciar la justificación teórica de la vida artificial, pero también existen muchas aplicaciones prácticas inmediatas, entre las que resaltan la animación realista por computador que se emplea ya habitualmente para simular estampidas de animales, bandadas de aves, cardúmenes de peces, fuego y un largo etcétera.

Sirve para entender la interacción entre seres vivos, sean plantas, animales o personas.

Sirve para simular otros mundos, en particular mundos de la física (autómatas celulares, velocidad de la luz, contracción del tiempo...) e incluso nuestro propio universo.

Sirve como un paso intermedio para lograr inteligencia artificial.

Y, por primera vez en la historia, sirve para hacer filosofía en el laboratorio. Aunque decirlo pueda escandalizar a muchos filósofos. La filosofía y, más concretamente, la metafísica en la Grecia clásica fue lo que propició el desarrollo posterior de la Física. Mirad que cosa tan increíble ocurrió hace 2400 años: cuando todavía no había instrumentos para medir casi nada, Demócrito pudo deducir que la materia estaba compuesta por átomos, aunque su definición de átomo fuera distinta a la nuestra y algunas de las razones que esgrimía estaban equivocadas, pero aun así fue un impulsor de la Física. Sin embargo, la Física nació definitivamente con Galileo, y entonces la filosofía se quedó completamente atrás, concretamente la metafísica, que se convirtió en un enredo sin sentido donde las palabras se sostienen por sí solas sin ningún apoyo de la realidad³ y es asaltada finalmente por la gente supersticiosa, las pseudociencias y los embaucadores.

La idea es la siguiente: si tienes una hipótesis filosófica nueva de algo para lo que es difícil montar experimentos en el mundo real, puedes crear una simulación en el computador y ver como se desenvuelve. Esto es filosofía práctica. Uno de los mejores filósofos que usa este método es Daniel Dennett, del que hablaremos extensamente más adelante.

MUCHOS MUNDOS

La mayoría de las obras artísticas que más nos llaman la atención contienen mundos virtuales que se “enredan” con el mundo real, hasta tal punto que se hace difícil distinguir uno de otro. A veces también aparecen bucles de autorreferencia, es decir, en un mundo X se crea un mundo Z, y en el mundo Z se habla o se modifica X. Como veremos más adelante, las autorreferencias son creativas, incluso las que generan contradicciones.

Ejemplos de ello son *Don Quijote de la Mancha* de Cervantes, formado por dos libros. Al principio del segundo libro, el Quijote habla sobre sus aventuras publicadas en el primer libro y relata las inexactitudes y errores cometidos por el escritor.

También *La historia interminable* de Michael Ende, formada por dos cuentos que se entrelazan (y que el editor marcó con letras en dos colores distintos, para facilitar la comprensión de los niños a los que va dirigido).

3 Ver, por ejemplo, Zubiri (1996) y Marchand (2015).

Los libros de Jostein Gaarder, como *El mundo de Sofía*, donde se desarrolla la historia de una niña real en paralelo a una niña de ficción, y donde finalmente el personaje de ficción escapa de su mundo virtual para irse a vivir al real. O *Maya*, donde hay historias entremezcladas y contradictorias (un personaje muere, y más adelante reaparece vivo). La contradicción se disuelve cuando, al final, nos explican que una de las historias es un libro escrito por otro de los personajes.

En el cine, *La rosa púrpura del Cairo* de Woody Allen, también con personajes de una película que escapan para vivir en la vida real, aunque, inevitablemente, siguen estando detrás de la pantalla.

En la pintura, muchas de las obras de Escher muestran un espacio plano, un dibujo 2D dentro del dibujo, entremezclado con un espacio 3D (dibujado también, por supuesto, pero en perspectiva), es decir, dos mundos distintos, uno con mayor apariencia de realidad que el otro (aunque obviamente ambos son virtuales), a veces en cooperación, a veces en contradicción.

Douglas Hofstadter, en su libro *Gödel, Escher, Bach, un eterno y grácil bucle*, se recrea en mostrarnos estos aspectos en las obras musicales de Bach; y en las pinturas de Magritte donde aparece un cuadro pintado dentro del cuadro real, o una pipa con un cartel que dice “esto no es una pipa”; o en los dibujos de Escher como una mano dibujando otra mano, que a su vez dibuja a la primera; o en la autorreferencia de los lenguajes, que da lugar al teorema de Gödel, que veremos en el capítulo “Computación Completa”.

Julio Cortázar, en muchos de sus cuentos. Basta señalar “La noche boca arriba”, donde un motorista sufre un accidente y va a entrar al quirófano de un hospital. A veces cae en delirios donde sueña que es un guerrero de un pueblo indígena, que los incas acaban de capturar y que van a sacrificar. Cuando despierta va avanzando la historia del enfermo, cuando duerme va avanzando la historia del indígena. O será al revés, nos dice Cortázar a la mitad del cuento —será que es un indígena que cuando duerme sueña que va cabalgando en un caballo de hierro—, dejándonos en la duda que nunca resuelve, pues al final ambos personajes mueren simultáneamente.



Fuente: CC BY 2.5, Maurizio Codogno (2002). Disponible en: <https://goo.gl/UHf9m9>

Personaje 1

Douglas Hofstadter (1945-)

Douglas Hofstadter es un científico de USA que trabaja en ciencias cognitivas, informática y filosofía, entre otras cosas. Entre sus publicaciones destacan *The Mind's I: Fantasies and Reflections on Self and Soul* (1981) en colaboración con Daniel Dennett, *Metamagical Themas* (1985) y *Gödel, Escher, Bach: un eterno y grácil bucle* (1989). En el primer libro recrea temas de inteligencia artificial y vida artificial usando cuentos de Stanislaw Lem y otros escritores, como hilo conductor. En el segundo libro habla de múltiples temas, entre los cuales están los fractales y el caos. Pero es el tercer libro el que lo catapulta a la fama: recogiendo las semillas plantadas en los libros anteriores muestra que la autorreferencia (un bucle de realimentación) es un proceso creativo capaz de llevarnos a la inteligencia y a la consciencia.

Las películas de cine *Abre los ojos* y *Los otros* de Alejandro Amenábar presentan un mundo donde poco a poco aparecen situaciones extrañísimas, paradójicas o contradictorias. Hacia el final se nos hace caer en cuenta que el personaje está viviendo en otra realidad donde todo es posible.

Otro escritor excepcional abordando el tema de “qué es la realidad” es Philip K. Dick. De varios de sus libros se han hecho películas de cine. Por ejemplo, *Ubik* presenta un futuro donde los muertos se pueden mantener en hibernación y conseguir de ellos alguna actividad cerebral (ensoñaciones) que les permiten vivir otra vida e incluso comunicarse con los habitantes del mundo real. En “Usted lo recordará perfectamente” llevado a cine como *Total recall* nos presenta un mundo donde viajar es demasiado caro. En su lugar, las agencias de viajes implantan en el cerebro del turista los recuerdos del viaje que deseen, para que pueda disfrutar de ello el resto del año. El personaje principal se da cuenta que su vida es solo una falsificación, los recuerdos de toda su vida han sido borrados y sustituidos por otros. Luchando por obtener su personalidad real averigua que es un agente secreto. Al final del libro comprende que esa tampoco es su verdadera vida, y comienza a hacer varias regresiones hacia vidas supuestamente reales pero que resultan no serlo, y cada una es más disparatada que la anterior. La mayoría de los cuentos de Dick donde presenta alternativas a la realidad se han llevado a cine: *Blade Runner*, *Minority Report*, *Paycheck*, *A Scanner Darkly*, etc.

También el *Congreso de Futurología*, de Stanislaw Lem, donde se muestra un futuro en el que la humanidad vive una vida semirreal semiimaginaria gracias al uso extendido de una multitud de drogas que alteran la percepción (es más barato comprar la droga que sirve para percibir que estoy viendo la televisión, a comprar el propio televisor). Estas drogas son usadas tanto por los propios individuos, para mejorar su bienestar, e igualmente son difundidas en el aire y agua por su gobierno, para manipular a la población. El final es también una regresión hacia atrás, en la búsqueda de la verdadera realidad del mundo, y dejando al lector la duda de si realmente se ha alcanzado.

Las películas *Piso 13*, *Matrix* y *Avalon* son las que plantean más directamente el asunto de mundos artificiales, mundos construidos dentro de un computador y donde se puede pasear e interactuar de una manera virtual aunque con consecuencias reales. La película *Inception* nos muestra otra posibilidad: que los mundos virtuales sean ensoñaciones de nuestro propio cerebro, donde puedan acceder otros personajes, pueda haber varios niveles, y que al final no sepamos cuál es el nivel que corresponde a la realidad. Y la última que conozco, pues seguro que hay más, es la serie de televisión *Westworld* de 2016, donde se recrea un ambiente del *far west* de hace más de 100 años, en un pueblo real con personajes aparentemente reales, pero que son robots, y donde los turistas pagan por vivir experiencias emocionantes como luchar contra pistoleros, aunque no pueden sufrir ningún daño porque los personajes están programados para ello.

MUNDOS REALES Y MUNDOS VIRTUALES

Después de ver todos estos provocativos ejemplos de la literatura y el cine, se puede constatar que la opinión generalizada (e ingenua) es que solo hay un mundo real, y que puede haber muchos mundos virtuales (historias escritas, pinturas, música, cine...). Incluso los sueños que uno tiene cuando duerme son también mundos virtuales.

En seguida pasaré a argumentar por qué es ingenua. De momento uno estaría interesado en clasificar los mundos de mayor a menor grado de “realidad”, es decir, el mundo real sería el primero, y después vendrían mundos virtuales con un alto grado de verosimilitud con el real (mucho realismo), y después, sucesivamente, irían los mundos más groseros, donde los detalles no estén cuidados y sea obvio para todos que se trate de mundos artificiales.

Por ejemplo, esa clasificación podría quedar así:

- Mundo real
- Sueños
- Películas de cine

- Libros
- Mundos artificiales generados por computador

Este es solo un ejemplo. No pretendo que esta sea una clasificación definitiva, ya que los mundos generados por computador y presentados con ayuda de aparatos especiales (gafas, guantes, etc.) están llegando ya a un grado de verosimilitud muy alto.

Pero también suele ser comúnmente aceptado que hay muchos sueños asombrosamente reales, en los que uno disfruta o sufre y no es consciente de que se trata simplemente de un sueño, excepto en el momento final en que se despierta. Se llaman sueños vívidos⁴ y hay incluso técnicas para lograrlos, incluyendo aplicaciones para teléfonos móviles celulares, así como grupos de discusión sobre ello en Internet.

También ocurre algo similar con las buenas películas de cine, en las que nos sumergimos tanto en la acción virtual que participamos de los sentimientos de los personajes.

Al leer un buen libro ocurre algo ligeramente distinto: el escenario y los personajes no se nos aparecen a los ojos, sino directamente a la mente, es decir, los recreamos nosotros mismos, siguiendo nuestros gustos y fantasías, y eso añade un grado de personalización a este tipo de mundos virtuales.

¿Puede haber un mundo con más visos de realidad que el propio mundo real? Pues, parcialmente sí. Hay películas que se proyectan en los cines 3D que son hiperreales, en el sentido de que ofrecen a la vista muchos más detalles de los que suele tener el mundo real. Pero digo “parcialmente” porque la vista no es el único sentido que participa de la percepción de realidad; el tacto, el gusto, etc., todavía no nos engañan en estos casos. Sin embargo, es solo cuestión de tiempo, porque la tecnología produce resultados espectaculares cada vez con mayor rapidez. Con ayuda del computador y dispositivos de realidad virtual (gafas, guantes, etc.) será posible incluso detectar las acciones musculares que toma el espectador y cambiar la película de acuerdo a ellas, modificando el ángulo de visión o incluso el desarrollo de la acción. Eso hoy día ya se está haciendo, aunque todavía falta mucho por lograr.

Entonces, es de esperar que en algún momento podamos ver una película de cine a la vez que la tocamos, olemos, etc., gracias a que en el medio impreso de la película están grabadas las informaciones correspondientes a todos los sentidos, y no solo al visual. ¿Podríamos decir, entonces, que ese mundo es más real que el mundo real?

4 Lucid dreams.

De nuevo sí, pero parcialmente. Porque yo me puedo distraer de la película y abandonarla. Yo me puedo quitar las gafas y los guantes virtuales. De hecho, durante toda la experiencia, esas gafas y guantes los siento como algo añadido (siento el peso de las gafas y el sofoco del guante) que no debería estar allí. Y que, efectivamente, en el mundo real no están allí.

Para evitar que el espectador perciba estas imperfecciones, lo ideal sería presentar la información sensorial pero no a los sentidos sino directamente al cerebro. Este tema no es nuevo en filosofía. Por ejemplo, quisiera detenerme en el libro *La conciencia explicada*, donde Dennett lo aborda (brevemente) y lo resuelve (aunque insatisfactoriamente).

Dennett habla del problema del “cerebro en un tarro”: un pobre cerebro separado de su correspondiente cuerpo, y que ha sido sumergido en un líquido que le proporciona los nutrientes adecuados para vivir. Así mismo, todos los nervios que habitualmente salían por la médula espinal han sido conectados a un computador que puede registrar los impulsos eléctricos provenientes del cerebro y, a su vez, generar e inyectar hacia el cerebro nuevos impulsos. El computador estaría así en disposición de simular el mundo real pero de una manera que hace al cerebro impotente para detectar el engaño. ¿O no?

Dennett argumenta que esto no es así: el cerebro se daría cuenta del engaño por varias razones, la principal de las cuales es que la información sensorial que el cerebro recibe del mundo real (que se puede medir como un “ancho de banda”, en bits por segundo) es muy rica y compleja, comparada con la simulación que puede producir un computador.

A este argumento se puede replicar de tres formas simultáneamente:

- a. Los computadores (especialmente las redes de computadores) de hoy ya son bastante potentes como para generar mundos muy complejos. Y si esa complejidad no fuera todavía suficiente, es solo una cuestión de tiempo alcanzarla.
- b. Se puede usar otro argumento de Dennett, en el mismo libro. Resulta que el ancho de banda de los sentidos es mucho menor de lo que aparenta ser: la mayor parte de la información que creemos ver (con la vista) realmente es solo una reconstrucción que hace el propio cerebro. Dennett pone un ejemplo sorprendente de un computador que presenta un texto en la pantalla y que hace bailar todas las letras excepto unas pocas en donde está concentrada la vista del humano. El humano (el propio Dennett) percibe la pantalla como estática, a pesar de que casi toda está bailando, porque los ojos solo capturan información del centro del punto de visión, y el cerebro reconstruye el resto a partir de imágenes anteriores.

- c. Si una persona nace así, con su cerebro dentro de un tarro conectado a un computador, entonces no tiene forma de realizar comparaciones respecto al “mundo real”. Por tanto, aceptará el mundo que percibe como el real, así sea de muy baja calidad, así contenga imperfecciones⁵. No puede compararlo con nada más. Es algo similar a lo que ocurre con los ciegos de nacimiento: no hay forma de que puedan entender o imaginar qué es el color. Este es, por cierto, el argumento de la película *Matrix*.

La única prueba ampliamente aceptada de que estoy en el mundo real es pellizcarme. Si siento dolor, es el mundo real. Espero que el lector entienda la fragilidad de esta prueba. En un sueño muy real, también podría sentir dolor. En una simulación computacional de la realidad, un cerebro en un tarro emitiría impulsos eléctricos adecuados para mover los dedos y comandar la acción de pellizcar el otro brazo. Estos impulsos los recibiría el computador y generaría como respuesta los impulsos eléctricos adecuados que emitiría un brazo real como señal de dolor. El cerebro recibiría una señal de dolor en el brazo, de donde deduciría, erróneamente, que el mundo en el que vive es el real.

Especialmente en la situación mencionada en c), no hay prueba que permita distinguir el mundo real de los virtuales, salvo la fe ciega de que vivimos en él. Y eso no es muy científico. Durante mucho tiempo esta cuestión fue abordada de forma tímida por la filosofía, dado que no había simulaciones de computador. Pero actualmente, con la proliferación de mundos virtuales cada vez más detallados, es una pregunta pertinente, cuya respuesta es contundente: si no hay forma de diferenciar lo real de lo virtual, es que no son diferentes.

RESUMEN

La vida artificial es una disciplina de las ciencias de la computación que intenta entender la complejidad del mundo, especialmente la de los seres vivos, por ser donde más complejidad hay.

5 Quizás contenga imperfecciones muy grandes como dejar un electrón confinado en un pozo de potencial y que después de un rato aparezca al otro lado. Allí tendrá dos opciones: decir que ocurrió algo mágico, o tratar de modelar ese tipo de errores llamándolos, por poner un nombre gracioso, “efecto túnel”.

En esencia no hay diferencia entre vida natural y vida artificial, ni entre mundos reales y mundos artificiales. Lo único que los separa es el grado de complejidad.

Para saber más

¿Quién no ha visto la película *Matrix*? Es un buen abrebotas a este tema. Pero mejor aún es *Piso 13*, al menos en lo referente a este libro. Todas las demás películas tienen algo sorprendente para aportar, así como los cuentos de ciencia ficción. La ciencia ficción es la nueva inspiración de la ciencia, y se realimentan mutuamente.

Una vez que queramos profundizar en estos temas, recomiendo leer los libros de Hofstadter y Dennett, con paciencia, pues aunque son muy amenos también son muy extensos.

No pierdan el tiempo ni con Zubiri, ni con Marchand. Están aquí como contraejemplos.

REFERENCIAS

Libros, artículos y enlaces web

- Bostrom, N. (2003). Are You Living in a Computer Simulation? *Philosophical Quarterly*, 53(211), pp. 243-255. DOI: https://doi.org/10.1111/1467_9213.00309 JSTOR 3542867
- Cervantes, M. (1998). *Don Quijote de la Mancha*. Madrid: Edimat Libros.
- Cortázar, J. (1994). La noche boca arriba. En *Cuentos completos 1*. Madrid: Alfaguara.
- Dennet, D. (1995). *La conciencia explicada: una teoría interdisciplinaria*. Barcelona: Ediciones Paidós.
- Dick, P. K. (1969). *Ubik*. Barcelona: Ediciones Orbis.
- _____. (1973). Usted lo recordará perfectamente. En *Ciencia Ficción 4*. Barcelona: Editorial Bruguera.
- Ende, M. (1983). *La historia interminable*. Madrid: Alfaguara.
- Ernst, B. (1978). *El espejo mágico de M. C. Escher*. Alemania: Taschen.
- Gaarder, J. (1997). *El mundo de Sofía*. Bogotá: Editorial Norma.
- _____. (2000). *Maya*. Madrid: Siruela.
- Hofstadter, D. (1979). *Gödel, Escher, Bach: un eterno y grácil bucle*. Barcelona: Tusquets Editores.
- _____. (1985). *Metamagical Themas*. New York: Basic Books.

_____. y Dennett, D. (1981). *The Mind's I: Fantasies and Reflections on Self and Soul*. New York: Basic Books.

Lem, S. (1981). *Congreso de Futurología*. Barcelona: Editorial Bruguera.

Marchand, A. (2015). *The Universe is Virtual: discover the science of the future, where the emerging field of digital physics meets consciousness, reincarnation, oneness, and quantum forgiveness*. United States: Inspired Arts Press.

Zubiri, X. (1996). *Espacio, tiempo, materia*. Madrid: Alianza Editorial.

Películas y videos

Allen, W. (1985). *La rosa púrpura de El Cairo*. Estados Unidos: Orion Pictures.

Amenábar, A. (1997). *Abre los ojos*. España, Francia e Italia: Sogecine.

_____. (2001). *Los otros*. España, Francia, Estados Unidos: Sogecine, Cruise-Wagner Productions y Las Producciones del Escorpión.

Hermanas Wachowski (1999). *Matrix*. Estados Unidos: Warner Bros.

Jansen, T. (2007). *My creations, a new form of life*. TED. Recuperado el 17 de agosto de 2017. Disponible en: <https://goo.gl/HeqNio>

Linklater, R. (2006). *A Scanner Darkly*. Estados Unidos: Thousand Words, Section Eight Productions, Detour Filmproduction y 3 Arts Entertainment.

Nolan, Ch. (2010). *Inception*. Estados Unidos: Legendary Pictures y Syncopy Films.

Nolan, J. y Joy, L. (2016). *Westworld*. USA: Bad Robot Productions, Kilter Films, Warner Bros. Television, Jerry Weintraub Productions.

Oshii, M. (2001). *Avalon*. Japón, Polonia: Nippon Herald Films.

Rusnak, J. (1999). *Piso 13*. Alemania y Estados Unidos: Columbia Pictures.

Scott, R. (1982). *Blade Runner*. Estados Unidos: Blade Runner Partnership.

Spielberg, S. (2002). *Minority Report*. Estados Unidos: Amblin Entertainment y Cruise/Wagner Productions.

Woo, J. (2002). *Paycheck*. Estados Unidos: Davis Entertainment y Lion Rock Entertainment.

COMPLEJIDAD

Para hacer una cosa, antes hay que hacer otra,

CLARA E. GARCÍA BAÑOS

La tesis que se va a sustentar a lo largo del libro es que no hay una diferencia radical entre mundos virtuales y el mundo real, entre objetos libres o con el futuro determinado, entre seres vivos e inertes, entre seres inteligentes y no inteligentes, entre seres conscientes o inconscientes, etcétera. La única diferencia es de grado (Figura 2), habiendo una amplia gama de objetos más o menos reales, más o menos inteligentes y más o menos libres. Y la forma más razonable y posiblemente la única de medir ese grado es usando la complejidad del objeto que estemos analizando.

El estudio de la complejidad es una rama de las matemáticas más bien reciente, pues tradicionalmente la ciencia ha trabajado en los dos extremos de una vasta escala:

- Si hay pocos objetos con pocas interacciones simples y lineales, entonces es relativamente fácil descubrir los comportamientos que se repiten, y ello da lugar a las teorías de la Física, como la mecánica clásica, el electromagnetismo y la relatividad.
- En el otro extremo están los fenómenos en los que intervienen billones de objetos con interacciones simples y lineales. En este caso es imposible saber en detalle qué es lo que hace cada objeto, pero la linealidad de las interacciones permite sacar promedios. Y ello es lo que da lugar a las teorías estadísticas, como la termodinámica y, en cierto modo, la mecánica cuántica.



Figura 2. Desde los sistemas menos reales hasta los más reales: cómic, cuento, película, ciudad. Desde los menos vivos hasta los más vivos: juego, fósil, orquídeas, papagayo. Y desde los menos inteligentes hasta los más inteligentes: herramientas, iguana, delfín, mico, Matías

Pero hay un inmenso territorio entre medias. Si lanzo al aire una piedra describirá una parábola que se puede calcular fácilmente con las ecuaciones de Galileo y Newton. Pero si lanzo al aire un pájaro, la trayectoria que siga es bastante impredecible, aunque al final también llegará a tierra. Incluso algo tan sencillo como la teoría de la gravedad de Newton, dentro de la mecánica clásica, solo es aplicable para interacciones entre dos cuerpos. La teoría no permite predecir en detalle qué va a pasar si hay tres o más cuerpos porque las ecuaciones de Newton no tienen solución analítica en este caso, únicamente se pueden hacer aproximaciones, y los errores se acumulan de forma descontrolada (lo vimos en el capítulo sobre el caos del libro anterior). Algunos de estos sistemas son analizados por las “ciencias de la complejidad” (otra manera de llamar a la “vida artificial”).

Podemos discutir sobre la diferencia entre vida artificial y vida real, inteligencia artificial e inteligencia natural, o entre computación real y computación simulada. Comencemos por la última: si entiendes la equivalencia entre todas las máquinas de Turing universales, entonces sabes que no hay ninguna diferencia. Un computador con un programa escrito en lenguaje máquina ejecuta esencialmente los mismos algoritmos que otro programado en alto nivel compilado, o dentro de un intérprete, o dentro de un simulador, o dentro de una máquina virtual como *VirtualBox OSE*, o dentro de mil máquinas virtuales, una dentro de otra. La única diferencia es la velocidad (e incluso así no se puede medir desde dentro del propio sistema, sino desde el nivel de la simulación anterior), pero el cómputo es el mismo en todos los niveles.

¿Cuál es la diferencia entre la inteligencia real y la simulada? Muchos filósofos tradicionalistas como Searle (1997) establecen que son cosas bastante distintas. Pero te propongo el siguiente juego, para que entiendas que no hay tal diferencia: supón que vas a clase de Cálculo I, y que eres muy malo en esos temas. No entiendes nada. El día del examen se te ocurre una idea salvadora. Dado que no sabes nada de cálculo, lo único que puedes hacer es fingir que sí sabes. De esta manera, finges muy bien y respondes correctamente a todas las preguntas del examen. Has logrado engañar al profesor, fingiendo que sabes, cuando realmente no sabes nada.

Espero haber despertado al menos una sonrisa en este tipo de razonamiento porque, efectivamente, eso no se puede hacer: si logras fingir que sabes cálculo, es porque efectivamente sabes cálculo. Y lo mismo pasa con la inteligencia.

Y lo que se puede decir de los mundos virtuales y el mundo real es lo mismo, o sea, que no hay ninguna diferencia esencial. En este sentido, nuestra querida realidad queda cuestionada. No sabemos si vivimos en una simulación realizada en algún tipo de computador. El filósofo Nick Bostrom (2016) hace un análisis de la cuestión y llega a la conclusión de que lo más probable es que nuestro mundo sea una simulación. Concretamente Bostrom dice que al menos una de las tres siguientes proposiciones tiene que ser verdadera:

- Prácticamente toda civilización que alcance nuestro nivel actual de sofisticación tecnológica, se extinguirá.
- La inmensa mayoría de las civilizaciones que alcancen una tecnología sofisticada no querrán ejecutar simulaciones de su propia historia o variantes de ella.
- Es muy probable que vivamos dentro de una simulación.

El argumento se basa en que si las proposiciones primera y segunda son falsas (lo cual suena muy razonable), entonces habrá un crecimiento exponencial de civilizaciones simuladas que ejecuten simulaciones de otras civilizaciones, que ejecuten simulaciones... Y eso implica que habrá un número enorme de civilizaciones simuladas, frente a un reducido número de civilizaciones reales (solo una). La probabilidad de pertenecer a una simulada es abrumadoramente alta.



Fuente: CC BY-SA 4.0, Future of Humanity Institute (2014). Disponible en: <https://goo.gl/eHhSsh>

Personaje 2

Nick Bostrom (1973-)

Niklas Boström es un filósofo sueco con ideas bastante atrevidas, muy alejadas de la filosofía tradicional. Fue fundador de la organización Humanity+, que busca promover el transhumanismo. En particular recomiendo dos temas que él ha desarrollado: el de que vivimos en una simulación, que acabamos de enunciar. Y el peligro de la Inteligencia Artificial, que se comentará en el capítulo del mismo nombre, que no tiene nada que ver con rebeliones de robots como nos presentan en muchas películas. Las inteligencias artificiales pueden hacernos mucho daño incluso empleadas con las mejores intenciones. En TED hay un video suyo en 2017 explicando estos peligros.

Obviamente es una idea con muchas controversias, entre otras, que no es falsable, es decir, que no se puede hacer ningún experimento para rebatirla. Pero el punto que quiero traer aquí es precisamente ese. No hay forma de distinguir un universo real de uno simulado, para los observadores que vivan dentro.

La otra crítica que quiero hacer aquí a esta argumentación es que, como veremos en un capítulo posterior, los simuladores tienen un tiempo interno que se ejecuta más lentamente que el tiempo externo real, pues para hacer una simulación hay que realizar cálculos que demoran bastante. Los simuladores se usan con dos propósitos: a) recreativos, es decir, para crear escenarios reales o ficticios donde interactuar en tiempo real, pero manteniéndose a salvo de consecuencias desagradables (si mueres en el juego, no mueres en la vida real); y b) investigativos, es decir, para recrear situaciones posibles del mundo real y observar cuáles serán sus resultados. En el primer caso se necesita que el universo simulado vaya igual de rápido que el real. En el segundo caso, se requiere que vaya incluso más rápido. De este modo, si

se avecina una catástrofe, la puedo observar en el universo simulado antes de que ocurra en el real, y puedo ensayar en el universo simulado formas de prevenirla, llevando a la práctica la alternativa que funcione mejor. Una simulación que vaya más lenta que el mundo real no le sirve a nadie, y por eso es que se necesita acelerar el tiempo del universo simulado. La única forma que existe de lograrlo es renunciando a simular los detalles. Así, los algoritmos del mundo simulado son más sencillos y se ejecutan más deprisa, pero a cambio se pierde la complejidad y riqueza que añaden los pequeños detalles del mundo (por ejemplo, simular cómo se va a comportar un gas es muy rápido de hacer si se usa la ecuación de los gases perfectos, pero esa ecuación nada nos dice de la existencia de los átomos).

A lo que quiero llegar es que si hay un simulador dentro de otro y este dentro de otro en un número indefinido de niveles, cada nivel tendrá menos detalles que el anterior, y eso impone un límite práctico a la simulación recursiva de mundos.

Como conclusión, podemos decir que es imposible el escenario que plantea Bostrom. No hay una cantidad exponencialmente grande de mundos simulados. A lo sumo puede haber una cantidad pequeña y finita de mundos simulados respecto al mundo real. ¿Significa eso que estamos en el mundo real? Tampoco. Hay una cierta probabilidad de que estemos en una simulación, pero no es la certidumbre que plantea Bostrom. Lo único que se pretende remarcar en este libro es que no hay forma de diferenciar un mundo real de un mundo simulado. Son equivalentes.

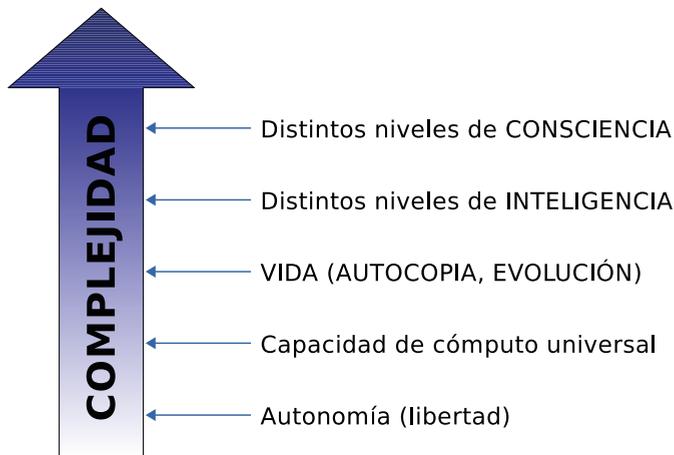


Figura 3. Al aumentar la complejidad aparecen nuevas características en un sistema

Pensemos ahora en qué es lo que hace falta para poder fabricar un universo simulado. Deben existir al menos capacidad de cómputo universal e inteligencia. ¿Cómo lograr estas cosas? A lo largo del libro veremos que es muy fácil, prácticamente inevitable. Resulta que conforme la complejidad de un sistema aumenta emergen nuevas características, tales como la capacidad para calcular, la libertad para tomar decisiones, la capacidad de cómputo universal, la autoorganización (entendida como la capacidad de defender la propia frontera⁶ y fabricar la estructura interna) la posibilidad de sacar copias de sí mismo, la inteligencia y la consciencia (Figura 3). Estas características no tienen por qué surgir exactamente en ese orden pues como argumentaré en los correspondientes capítulos, la inteligencia y la consciencia tienen muchas facetas y algunas pueden aparecer de forma temprana y otras más tarde. Además, puede haber más características que no estén aquí identificadas o incluso que surjan más adelante en la historia del universo.

Todas estas características son graduales, a la vez que pueden intensificarse, emergiendo nuevos matices y relacionarse de forma compleja unas con otras. Por ejemplo, cuando surge la consciencia ello implica un bucle de introspección, que puede mejorar o empeorar la libertad del individuo según predomine la realimentación positiva o la negativa.

De la misma manera, cuanto más inteligente es un agente más se esfuerza por aumentar su propia libertad, pues ello le permitirá acceder a un abanico más amplio de posibilidades, típicamente con el objetivo de sobrevivir. Ello conforma también un bucle de realimentación, en este caso, realimentación positiva. Pensemos, por ejemplo, en el juego del ajedrez. Si no sabes mucho ajedrez, una buena estrategia es aquella que aumenta el número de posiciones alcanzables; y una mala estrategia aquella que las disminuye. La peor jugada posible es en la que entregas tu rey para que lo elimine el adversario, ya que después no tendrás ninguna otra jugada disponible pues has perdido la partida. Y dado que nadie sabe cuál es la mejor jugada para un tablero dado, si aumentas los estados alcanzables es más probable que esa mejor jugada se encuentre allí. Mientras que si disminuyes los estados alcanzables, pierdes libertad, y tu contrincante podría analizar mejor lo que planeas hacer (ya que puedes hacer menos cosas) y encontrar una estrategia contra ti. En la vida real humana pasa lo mismo: aunque no puedas saber en detalle lo que te espera en la vida, estudiar una carrera suele ser una buena decisión porque aumenta tus posibilidades futuras, mientras que caer en una adicción suele ser malo porque disminuye el número de cosas que puedes hacer después. Basándose solo en esto, se ha desarrollado un algoritmo que

6 En la misma línea que la definición de autopoiesis dada por Maturana y Varela en 1973.

juega muy bien cualquier tipo de juego, incluso sin conocer las reglas, y que aprende progresivamente a base simplemente de aumentar sus posibilidades futuras (Wissner-Gross y Freer, 2013).

Pero también, cada una de estas propiedades requiere un mínimo de complejidad para que surjan, que iremos viendo en cada capítulo. Y ello establece una serie de condiciones necesarias, aunque no suficientes. Por ejemplo, si diseñas un sistema con 5 bits de complejidad, podemos estar seguros de que no tiene apenas libertad y definitivamente no le alcanza para tener computación completa. Mucho menos para ser inteligente o consciente.

Vamos a estudiar entonces cómo se puede medir la complejidad de un sistema, cómo se puede generar (podemos adelantar que el mecanismo más importante es la realimentación, que ya la hemos visto en el libro anterior) y, sobre todo, por qué se genera espontáneamente. Con ello se espera dar solución, así sea parcial, a la famosa, filosófica y fundamental pregunta de la ciencia: “¿Por qué hay algo en vez de nada?”⁷.

¿POR QUÉ SE GENERA COMPLEJIDAD DE FORMA ESPONTÁNEA?

La evolución es un mecanismo asombroso para generar complejidad, como sabe cualquiera que haya experimentado con los algoritmos evolutivos que vimos en el libro anterior.

Pero cuando un sistema casi no tiene nada, cuando aún no existe la reproducción (con lo cual la evolución todavía no se ha puesto en marcha), cuando quizás incluso no tenga capacidad de cómputo completa (aunque sí se requiere una cierta capacidad de cómputo), todavía entonces es muy posible que pueda hacer crecer su propia complejidad. Vamos a ver cómo. Para ello primero se dará una breve introducción a varios conceptos matemáticos (¿Qué esperabas? Aunque el sistema no tenga casi nada, las matemáticas siempre están allí). Y después se discutirá cómo ello sirve para generar complejidad.

No sabemos si estas formas de generar complejidad son las únicas, o incluso que sean varios aspectos de un único fenómeno, pero son las que se han identificado hasta ahora. Si las ordenamos de la más simple a la más exigente en requerimientos son:

- **Referencia y autorreferencia.** Un buen libro para aprender más sobre ello es el de Hofstadter (1979). Las autorreferencias (Figura 4) conforman el bucle de realimentación más sencillo, y podemos esperar

⁷ Hay ya muchos investigadores trabajando en ello con excelentes propuestas. Ver, por ejemplo, Atkins (1992) y Krauss (2009).

de ellas que produzcan algún fenómeno complejo. Pero también se comienza a estudiar los aspectos filosóficos de ello, en lo que tiene que ver con las referencias y autorreferencias de la materia (Pattee, 2012).

- **Combinatoria.** Agrupar varios objetos para crear otro más grande. Esto, aunque es trivial, se puede hacer de muchas maneras distintas, por lo que se genera una gran diversidad de nuevos objetos. A su vez, los objetos agrupados interactúan entre sí de varias maneras que se pueden estudiar por medio de la teoría de juegos y suelen crear bucles de realimentación.
- **Realimentación.** Que, a su vez, suele producir:
 - **Fractales.**
 - **Caos.**
 - **Leyes de potencias.**
- **Evolución.** Además, puede emerger la cooperación entre los objetos que evolucionan, y ello es un fenómeno muy potente que genera mucha más complejidad.

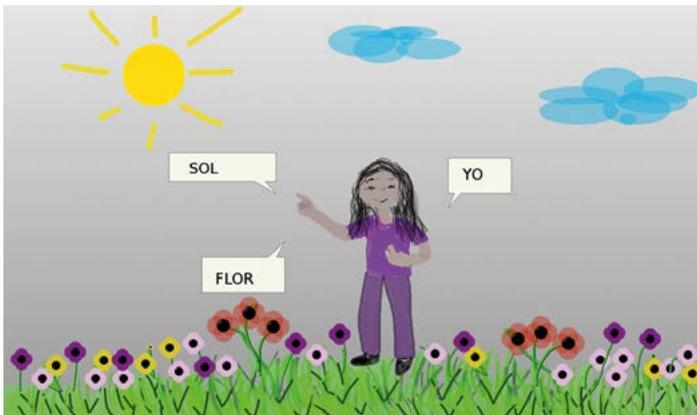


Figura 4. Referencias y autorreferencia en lenguaje hablado

Fuente: dibujo realizado y cedido amablemente por Helga R. Calvo R. (2017)

Vamos a recordar todo ello, aun cuando ya vimos los detalles en el libro anterior.

La referencia es algo tan sencillo como darle nombre a las cosas. En el lenguaje diario empleamos las palabras para referirnos a los objetos (Figura 4). En el computador empleamos los nombres de las variables para referirnos a los datos que contienen. Si pensamos en lo más básico del computador, el código máquina (o, para mayor facilidad, en lenguaje C en la figura 5), cada dato se guarda en una dirección de memoria que equivale a su nombre. De este modo, en una dirección de memoria se puede guardar un dato o

bien otra dirección de memoria que apunte a un dato, y esto último es lo que se llama una referencia o un puntero en lenguajes de alto nivel (no se usa directamente el dato, sino que nos referimos a él por medio de su dirección en memoria) o es un direccionamiento indirecto, en lenguaje ensamblador. También en una dirección de memoria se puede guardar ella misma (en lenguaje C eso sería un puntero apuntando a sí mismo) constituyendo una autorreferencia.

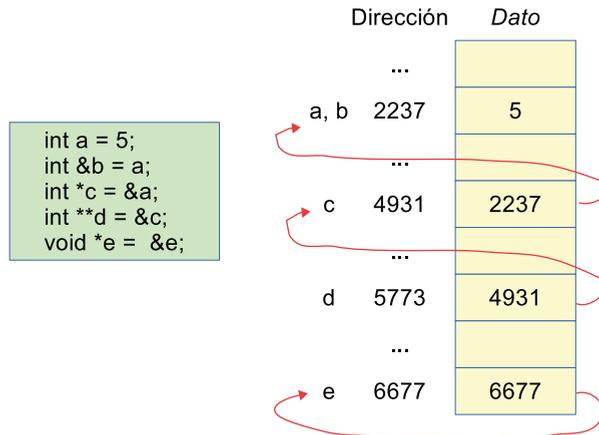


Figura 5. Referencia, puntero, puntero a puntero y puntero a sí mismo, en lenguaje C++ (las direcciones son arbitrarias)

En la figura 6 podemos ver más ejemplos de referencias un tanto insólitas: los avisos de las áreas en los laboratorios universitarios para cumplir con ciertos estándares de calidad, que a su vez nos recuerdan la historia relatada por Oliver Sacks en su libro *Un antropólogo en Marte*, donde visitó a una familia de autistas que tenían etiquetados todos los objetos de la casa.

Las referencias constituyen el paso previo para las autorreferencias, que son las que realmente generan complejidad, bien sea porque construyan recursiones infinitas como un fractal, bien sea porque conduzcan a contradicciones que destruyen el sistema, lo que obliga a crear uno nuevo para poder salir de ellas. Como ejemplo divertido del primer caso, me acuerdo de una carretera cerca de donde vivo que tiene una secuencia de señales de tráfico tradicionales (prohibido aparcar y límites de velocidad), pero la última de ellas es muy original y no está en ningún documento oficial, pues dice “respeten las señales de tráfico”. Supongo que lo lógico es añadir después otra que diga “respeten las señales de tráfico incluido las señales que dicen «respeten las señales de tráfico»”. Y luego otra, y otra, y así hasta el infinito.



Figura 6. Más ejemplos de referencias

Como ejemplo asombroso del segundo caso, recuerdo hace muchos años a mi pareja tratando de imprimir un dibujo por medio de un *software* muy mal hecho. Estaba tan mal hecho que aunque seleccionábamos la opción de imprimirlo horizontalmente, el *software* lo hacía en vertical, desperdiciando parte de la hoja a la vez que se perdían los laterales del dibujo. Después de muchos intentos descubrí, entre los menús enrevesados que tenía, que también se podía especificar esa opción por otro sitio (muy mal *software* porque cada opción debe de estar en un único lugar). Sin embargo, tampoco funcionó. En un momento de inspiración se me ocurrió especificar en un sitio “horizontal” y en el otro sitio “vertical” y ahí ocurrió algo memorable: apareció un letrero de error, regañándome y pidiendo que no me contradijera, a la vez que me preguntaba si lo que quería era imprimir horizontal o verticalmente (definitivamente un *software* pésimo porque tenía una funcionalidad ¡repetida tres veces!). Entonces seleccioné “horizontal” y ahora sí, lo imprimió como queríamos. Desde entonces pienso que las contradicciones profundas pueden ser una forma de aprendizaje, de despertar a la consciencia. El *software* despertó e imprimió correctamente. Y yo desperté y aprendí una nueva forma de aprendizaje. Mucho tiempo después también aprendí que los psicólogos llaman a esto “disonancia cognitiva” y, efectivamente, puede

llevar a las personas a que se replanteen ideas o comportamientos profundamente arraigados, pero erróneos, cuando acumulan demasiadas contradicciones en su contra.

Además, una vez que existen objetos referenciables puedes combinarlos para formar objetos más grandes, que es otra forma de generar complejidad. En el universo en que vivimos, si tienes quarks, electrones y fotones, la combinación de ellos forma átomos estables, cuya combinación forma moléculas, cuya combinación forma células, cuya combinación forma cuerpos, cuya combinación forma sociedades, dicho rápidamente y saltando algunos pasos y matices. Y no todas las combinaciones son posibles pues hay restricciones que lo impiden. Por ejemplo, restricciones geométricas, pues puede haber formas que no encajen. O la valencia de los átomos, que debe de ser compatible para poder formar moléculas. Y precisamente por ello hay complejidad, porque si no hubiera restricciones —si todo fuera posible— no haría falta analizar ni estudiar nada. No solo en nuestro universo sino en cualquier otro universo artificial que diseñemos, la variedad inicial también se dará por combinatoria.

Las realimentaciones producen dependencias. Las variables o los entes que conforman el sistema dejan de ser independientes y se influyen unos a otros. Eso puede producir realimentaciones positivas y caos, por ejemplo, si un pequeño cambio en una variable produce un cambio más grande en esa misma variable. Puede producir fractales si para fabricar una cosa primero hay que fabricar la misma cosa pero en una escala menor. Y esto es una realimentación en el algoritmo de fabricación, lo que se llama recursividad en la jerga computacional. O puede producir leyes de potencia como resultado de la misma realimentación positiva. Por ejemplo, cuanto más grande es una ciudad, más oportunidades ofrece y entonces atrae a más gente, haciéndose la ciudad aún más grande. Las realimentaciones también pueden producir libertad, pues, como Popper demostró, una máquina determinista no puede predecir su propio comportamiento ya que al conocer esa predicción tendría la alternativa de elegir hacer otra cosa distinta. Y, según Mackay, se puede predecir lo que va a hacer una persona o un grupo de personas, pero al revelarles la predicción se corre el riesgo de que reaccionen contra ella. Eso es lo que ocurre cuando se hace una encuesta electoral y se publica el resultado, pues la gente puede modificar su intención de voto. Y ello es imposible de evitar. No se pueden incluir los cálculos de la reacción que ocurrirá al conocerse el resultado. Revelar el pronóstico meteorológico a la gente no hace cambiar el tiempo, pero revelar el pronóstico económico sí hace cambiar la economía, porque el tiempo no depende de la gente, pero la economía sí, y se forma un bucle de realimentación (Barrow, 1999, pp. 338-342).

Los sistemas dinámicos caóticos, los fractales y las leyes de potencia generan mucha complejidad a partir de muy poco. Son procesos creativos. Y están ciertamente relacionados entre sí, pero no son iguales. Ha habido mucha confusión con este tema. En resumen se puede decir que:

- Prácticamente todos los fenómenos caóticos tienen como atractor un fractal. Se sabe que el atractor podría ser también un objeto de infinitas dimensiones, pero aunque sea una posibilidad matemática, no es muy realista. De modo que si hay caos, hay un fractal.
- A su vez los fractales pueden generar complejidad de tres formas: como atractores de sistemas dinámicos, que acabamos de comentar; como algoritmos para la construcción de estructuras espaciales, típicamente en plantas (Prusinkiewicz y Lindenmayer, 2004); o directamente como secuencias temporales, como propone Mandelbrot (2006) usando fractales aleatorios.
- Las leyes de potencias son distribuciones de probabilidad de eventos que no tienen un valor promedio definido y que decaen muy lentamente (lo que se llama “colas gordas de distribución”). Informalmente se dice que modelan sucesos probabilistas donde los eventos más grandes ocurren poco y los más pequeños ocurren mucho. Por otro lado, la relación entre leyes de potencias y fractales está en que ambos son autosimilares.

Conforme la complejidad aumenta, un objeto puede referirse a sí mismo en detalle, en cada una de sus partes, en la forma de construirse. Tenemos entonces el proceso de autocopiado, con lo cual también aparece la evolución, que genera mucha variedad y que habitualmente se percibe como un proceso de adaptación al entorno, lo que los ingenieros llamamos optimización.

Ya tenemos en nuestro sistema referencias, autorreferencias, agrupación combinatoria, realimentaciones y evolución. Pero también pueden producirse fenómenos donde intervengan simultáneamente varias de estas funcionalidades, generando mayores niveles de complejidad. Enseguida presentaré un algoritmo evolutivo generador de complejidad muy general, aunque podría haber otros.

COMPLEJIDAD VERSUS ENTROPÍA

Tenemos mecanismos que generan complejidad, pero ello parece contradecir la segunda ley de la termodinámica, que dice que en un sistema cerrado la entropía nunca disminuye sino que se mantiene o va en aumento, es decir⁸:

$$\frac{dS}{dt} \geq 0 \quad \text{Ec. 1}$$

Es importante entender que esta ley no dice absolutamente nada sobre los sistemas abiertos (que intercambian materia o energía con su exterior). En ellos la entropía puede aumentar o disminuir sin ningún problema.

Esta ley también se formula como que no es posible construir un *perpetuum mobile* de segunda especie, es decir, una máquina que funcione a partir de una única fuente de calor, o sea, que transforme calor en trabajo. En broma se dice que la primera ley de la termodinámica (de la conservación de la energía) expresa que no puedes ganar, mientras que la segunda dice que tampoco puedes empatar.

La entropía (S) tiene varias definiciones: por ejemplo, desde la física se refiere a una medida de lo próximo que está un sistema a un punto de equilibrio; desde la teoría de la información se habla de lo difícil que es predecir una secuencia de símbolos; y hay muchas otras alternativas. La definición de la física clásica se le debe a Boltzman y dice que la entropía de un macroestado depende del número (W) de microestados equivalentes que represente:

$$S = K_B \ln(W) \quad \text{Ec. 2}$$

Siendo K_B la constante de Boltzman.

A nosotros nos sirve entenderla como el grado de desorden de un sistema, aunque no sea una definición completamente exacta. Y, entonces, la segunda ley de la termodinámica dice que el desorden de un sistema cerrado siempre aumenta. Teniendo en cuenta que la vida tiene un componente importante de autoorganización, podría parecer que viola esta segunda ley. Sin embargo, no es así. Pensemos que la primera célula de un animal se va dividiendo y organizando en sus distintos órganos, hasta formar un cuerpo que nace, y que luego sigue creciendo a lo largo de su vida. Todo ello lo hace generando más orden, pero no es un sistema cerrado, pues necesita energía y alimento del exterior.

8 La entropía puede aumentar esporádicamente, pero la probabilidad de que lo haga es muy baja y decrece exponencialmente conforme aumenta el número de partículas involucradas. Por ello, en sistemas macroscópicos la probabilidad de aumento es cero a todos los efectos prácticos.



Figura 7. (a) Un lagarto es un sistema abierto, pues recibe comida y calor del exterior; (b) la Tierra con el Sol conforman un sistema razonablemente cerrado

Fuente: dibujos realizados y cedidos amablemente por Helga R. Calvo R. (2017)

Es decir, el ser vivo aumenta su orden (Figura 7-a), pero lo hace a expensas de desordenar considerablemente su entorno. Si luego consideramos todo el sistema ecológico, e incluso todo el planeta Tierra donde crece ese ser vivo, tampoco es un sistema cerrado, pues recibe energía del Sol. Pero si finalmente consideramos el conjunto Tierra+Sol, a efectos prácticos⁹ lo podemos aceptar como un sistema cerrado (Figura 7-b), y en él la cantidad total de desorden que crean los seres vivos durante su existencia compensa y supera el orden que autogeneran para sí mismos, de modo que allí sí se cumple la segunda ley de la termodinámica, como se supone que debe de ser.

Cuanto más desordenado, más entropía tiene. Pero ¿qué significa exactamente “orden” y “desorden”? No hay una respuesta sencilla, pero en los sistemas simples como los constituidos por partículas sin interacciones (gravitatorias, electromagnéticas, ni ninguna otra) hay más orden cuantas más simetrías puedas encontrar: en la figura 8, en (a) no hay simetrías, en (b) hay 8 simetrías (4 formas de rotar la figura para que quede igual más 4 simetrías especulares), y en (c) hay 12 formas (8 rotacionales y 4 especulares).

De hecho, la segunda ley de la termodinámica no es una ley física, sino matemática (de la estadística, concretamente). Se cumple en cualquier sistema, aunque sea construido dentro del computador con leyes físicas distintas. Y lo que viene a decir en su formulación matemática es que hay muchísimas más formas de desordenar un sistema que de ordenarlo, debido a que los sistemas con simetrías son pocos. De modo que si tenemos un sistema con una cierta entropía y lo observamos un tiempo después, lo más probable es que

⁹ Esto es, desde el punto de vista de la Tierra. Porque en realidad no es cerrado en absoluto: la mayor parte de la energía solar se pierde en el espacio, e incluso la Tierra pierde suavemente hidrógeno y helio, y gana diversos materiales que llegan en los meteoritos.

haya hecho una transición hacia un estado de mayor entropía, simplemente porque hay más estados de esos.

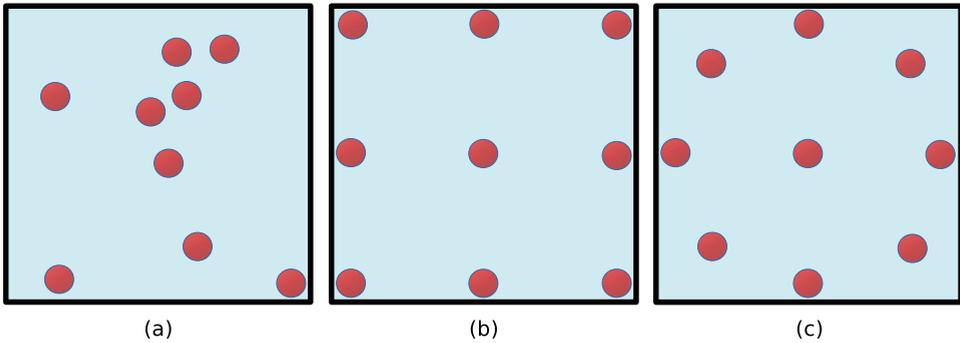


Figura 8. Estados: (a) desordenado; (b) ordenado; (c) muy ordenado

Si vertimos crema blanca en el café oscuro y revolvemos con una cuchara, al final se llegará a una mezcla homogénea de un color marrón claro; ese es un estado de equilibrio termodinámico, de máxima entropía, de máximo desorden. Y, como todo el mundo sabe, no es posible deshacer la mezcla moviendo la cuchara en sentido contrario. No es posible pasar del desorden al orden porque es muy poco probable que ello ocurra espontáneamente. Otro ejemplo que se suele poner es: si tienes una tienda de cerámica y entra un grupo de micos, después de un rato ¿cuál es el estado más probable? Obviamente la respuesta es: mucha cerámica rota por los suelos. ¿Hay alguna esperanza de volver a introducir los micos para que dejen todo como estaba? De nuevo, la respuesta es negativa. Para los físicos, el aumento de la entropía tiene mucho que ver con lo que se llama “la flecha del tiempo”, que siempre se mueve del pasado hacia el futuro.

Otro ejemplo más abstracto: si tenemos una baraja de 36 cartas que está completamente ordenada excepto porque están intercambiadas el 2 y el 3 de diamantes, diremos que tiene una entropía muy baja (un orden muy alto). Y si intercambiamos dos cartas elegidas al azar, la probabilidad de elegir precisamente las dos que están mal es bajísima (concretamente $p_{\text{ORDEN}} = 2 \cdot 1/36 \cdot 1/35 = 0.0016$), mientras que la probabilidad de que el cambio desordene todavía más la baraja es muy alto (concretamente $p_{\text{DESORDEN}} = 1 - p_{\text{ORDEN}} = 0.9984$). Inevitablemente la entropía aumentará conforme hagamos cambios al azar.

Entonces la vida no viola la segunda ley de la termodinámica, pero investigadores como el premio Nobel de Química Ilya Prigogine se han preguntado cómo logra disminuir localmente la entropía aun a costa de aumentarla globalmente en todo el sistema. Prigogine (1990, 1997) ha formulado con-

jeturas como que los sistemas tienden a generar más información (o más estructuras ordenadas) cuanto más lejos del equilibrio están, pero por desgracia eso no se cumple siempre, aunque sí en muchos casos. Conocemos muchas reglas (las llamadas leyes de la física) para, por ejemplo, convertir materia en energía. No obstante, todavía no tenemos una matemática de la autoorganización o, dicho con otras palabras, las reglas que convierten la energía en bits de información y viceversa, aunque ya hay algunos avances pues se ha resuelto una famosa paradoja de la termodinámica que ocurre en un proceso similar a mezclar café con crema.

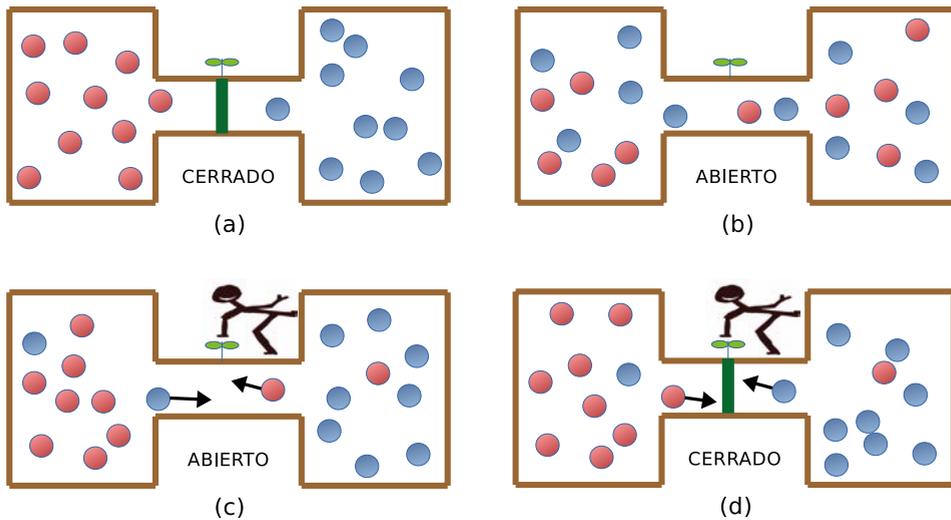


Figura 9. Entropía: (a) inicialmente baja; (b) hasta que se permite la mezcla; (c) el demonio de Maxwell disminuye la entropía al permitir el paso de ciertas moléculas; (d) mientras que impide el paso de otras

Para entender esto pensemos en dos habitaciones, la de la izquierda con un gas caliente, es decir, con moléculas cuya velocidad promedio es alta, y la derecha con el mismo gas pero a menor temperatura, con moléculas más lentas (Figura 9-a). Si abrimos la compuerta que separa ambas habitaciones los dos gases se irán mezclando, y al final la temperatura de ambas habitaciones será la misma, es decir, habremos alcanzado el equilibrio térmico maximizando así la entropía de este sistema (Figura 9-b). Sin embargo, una vez alcanzada esta situación, en 1867 a James Maxwell se le ocurrió que la compuerta se podría abrir y cerrar selectivamente. Concretamente la abriremos si detectamos que viene una molécula rápida desde el lado derecho, o una lenta desde el lado izquierdo, dejándolas pasar (Figura 9-c). En los demás casos la dejaremos cerrada (Figura 9-d). Con este truco podríamos volver a la situación inicial donde las moléculas rápidas están en el lado

izquierdo, sin invertir prácticamente nada de energía en el proceso. Al dispositivo capaz de detectar la velocidad de las moléculas para dejarlas pasar selectivamente se le denominó “demonio de Maxwell”.

Al disminuir la entropía tendríamos disponible energía gratuita. Y esto no puede ocurrir, según las leyes de la termodinámica. La paradoja se resolvió al entender que el demonio de Maxwell requería procesar información para realizar su tarea, concretamente detectar la velocidad de las moléculas que se estaban acercando y tomar decisiones con la compuerta, y se entendió que en los procesos reversibles, donde la cantidad de información permanece constante, la entropía también permanece constante. Solo si se destruye (se borra) la información ocurre aumento de entropía, esto es, en los procesos irreversibles. A partir de esto, Leó Szilárd en 1929 y posteriormente Rolf Landauer y Charles Bennett dedujeron una equivalencia entre información y energía o, más exactamente, entre entropía de la información y entropía termodinámica¹⁰: un bit de información equivale a 2.76×10^{-21} julios de energía. Volviendo al ejemplo de las dos habitaciones, el demonio de Maxwell sí puede separar las moléculas frías de las calientes, pero para hacerlo requiere mucha información, que equivale a una energía que hay que introducir del exterior, compensando la energía que se gana al separar las moléculas. La segunda ley de la termodinámica permanece entonces inviolada. Estos resultados tienen todavía bastante controversia pero también apuntan a que el mundo en que vivimos es computacional.

Desde otra perspectiva, John Barrow (1999) hace una observación muy interesante que logra incluir la información dentro de la Física: para que un sistema complejo pueda operar requiere tiempo, energía e información. Si le falta alguna de esas tres cosas puede compensarlo teniendo mucho de las otras dos. Por ejemplo, si tiene mucha información y tiempo puede crear energía, como vimos que hace el demonio de Maxwell. Si tiene mucha energía y tiempo, puede construir información, por prueba y error, creando sistemas autoorganizados como hace la evolución. Y si tiene mucha energía e

10 Todo esto sigue siendo un área activa de investigación. Por ejemplo, en el artículo “Entropy Production in Ecosystems” de Virgo y Harvey que se puede encontrar en Almeida (2007, p. 123) presentan un modelo minimal donde verificar el principio de máxima producción de entropía, mencionado anteriormente y defendido por Prigogine. Las restricciones en el sistema que presentan (una membrana que concentra un producto químico en su interior) hacen inviable la aparición de ciertos estados, lo que significa que disminuyen la entropía. Allí nos cuentan que la probabilidad de que se forme espontáneamente una bacteria *Escherichia Coli* es de 10^{-21} , lo cual generaría una entropía negativa de 3.2×10^{-12} J/K. La misma entropía se genera si 2.3×10^{-13} kg de agua se calientan de 300K a 301K. Este principio no es ampliamente aceptado pues no siempre funciona, pero modela muy bien las atmósferas de Tierra, Marte y Titán.

información, puede construir tiempo, en el sentido de saltar hacia el futuro, anticipando lo que va a ocurrir. Esto puede parecer confuso pero espero aclararlo con el siguiente ejemplo: si tienes mucho dinero (energía) y mucha información, puedes hacer la apuesta correcta en la bolsa de valores, con lo cual el futuro llega antes, al nivelarse los precios a sus valores correctos de mercado (en el momento en que usas tu información privilegiada para realizar una transacción, se hace pública) y, de paso, hacerte rico. Desde que se logró introducir el concepto de información en el mundo de la Física, ha aparecido una nueva forma de abordar esta última, llamada Física Digital¹¹.

¿CÓMO MEDIR LA COMPLEJIDAD DE UN SISTEMA?

Se puede medir con un único número y así lo vemos en la mayoría de publicaciones. Este número suele ser bits de información: la cantidad de bits que se requieren para describir un objeto, para transmitirlo, para producir los mismos resultados, etc.

Sin embargo, hay sistemas cuya complejidad requiere especificar al menos dos números. Por ejemplo, en el cerebro humano parece que lo importante no es solo su número de neuronas (del orden de 10^{11}) sino también el número típico de conexiones que salen de una neurona (del orden de 10^4). Las neuronas son un tipo de grafo, y en un grafo, efectivamente, hay que especificar el número de nodos y el número de arcos.

En un futuro deberíamos disponer de una teoría de la complejidad que sea capaz de predecir cuántos objetos y qué relaciones entre ellos se requieren para lograr la emergencia de un nuevo nivel de complejidad, así como cuáles son las relaciones requeridas entre esos objetos (en número y calidad). Pero ese momento aún no ha llegado. No parece que baste con uno o dos números para caracterizar la complejidad. Quizás se necesite un vector de números, una matriz u otra estructura de datos más sofisticada para describir la complejidad de un objeto. Es posible incluso que no haya forma de resumirla, que cada sistema complejo sea complejo a su propia manera, que sea único, y que la única forma de modelarlo sea especificando todos sus bucles de realimentación positivos y negativos. O quizás algún formalismo matemático no inventado aún. Pero de momento, como queda dicho, se mide con un único número, en bits.

11 *Digital Physics*.

Y ¿qué es un bit? Es la cantidad básica¹² de información. Su descubrimiento o definición se lo debemos a Claude Shannon (1948), que podemos ver, en términos sencillos en el recuadro 1.

Recuadro 1: Definición de bit

Un bit de información es una pregunta cuya respuesta es de tipo “sí” o “no” con ambas respuestas equiprobables.

Por ejemplo, si elegimos dos personas al azar en la calle y nos preguntamos “¿la primera persona lleva más monedas que la segunda?”, entonces la respuesta a esa pregunta conlleva un bit de información. Si averiguo la respuesta, sabré un poquito más acerca del mundo en que vivimos, concretamente tendré 1 bit más de información.

Un ejemplo incorrecto: tengo una nueva vecina y quisiera saber si sabe bailar salsa. La respuesta es “sí” o “no”, por lo que parecería que al hacer la pregunta voy a obtener un bit de información, pero se me olvidaba decir que estoy en la ciudad de Cali, donde casi todo el mundo practica este baile. Es mucho más probable que me diga “sí”, por lo que obtengo menos de un bit de información.

Vamos a dar una definición formal y general de bit. Para ello pensemos que va a ocurrir un evento X con n posibles resultados $\{r_1, r_2, \dots, r_n\}$ y sé que la probabilidad de obtener cada resultado es, respectivamente, $\{p_1, p_2, \dots, p_n\}$. Si espero a que ocurra el evento y averiguo el resultado, la información H que voy a obtener de ese evento X es:

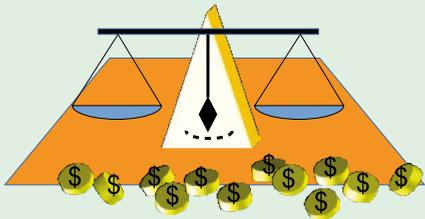
$$H(X) = \sum_{i=1}^n p_i * \log_2 \left(\frac{1}{p_i} \right) \quad \text{Ec. 3}$$

También se llama cantidad de información o entropía del evento X , y sus unidades son bits. Es una medida de la dispersión de la probabilidad $p(x_i)$: cuando hay mucha dispersión, obtengo mucha información cuando ocurre el evento. Y viceversa.

En el primer ejemplo de las monedas, como no tenemos ninguna pista acerca del resultado, las dos respuestas son equiprobables y entonces $p_1=p_2=1/2$, por lo que $H=1/2*\log_2(2)+1/2*\log_2(2) = 1 \text{ bit}$.

En el segundo ejemplo, si el 80% de l@s caleñ@s bailan salsa, entonces $p_1=80\%$ y $p_2=20\%$, por lo que la cantidad de información que obtendré al saber la respuesta es $H=80/100*\log_2(100/80)+20/100*\log_2(100/20) = 0.72 \text{ bits}$.

¹² No es la cantidad mínima porque puede haber fracciones de bit, como veremos en el problema 1.



Problema 1: Las doce monedas

Tenemos 12 monedas iguales, salvo que una de ellas es falsa. Y se sabe que es falsa porque pesa distinto que las demás (quizás pese más o quizás pese menos). Tenemos también una balanza de dos platillos, de esas antiguas, que indica únicamente cuál platillo pesa más o si ambos pesan igual. Usando la balanza solo tres veces, averigua cuál es la moneda falsa y di si pesa más o pesa menos que las otras.

Pista: la pregunta que hagas (forma de pesar) debe dar resultados equiprobables.

Esto significa que si tenemos un evento cuyo resultado es conocido con muy alta probabilidad, la información que nos aportará cuando ocurra es muy pequeña. Mientras que si todos los resultados son igual de probables entonces tendremos la máxima incertidumbre, de modo que cuando ocurra el suceso obtendremos la máxima información posible.

Un tercer ejemplo podría ser: en un aula de clase, pedir a alguien que piense en uno de los objetos que allí hay (tablero, proyector, computador, equipo de sonido, aire acondicionado, mesa del profesor, silla del profesor, puerta). Se trata de adivinar cuál objeto es, haciendo preguntas cuya respuesta sea sí o no, usando el menor número de preguntas posible. Para ello, la mejor estrategia es hacer preguntas cuya respuesta sea equiprobable. Por ejemplo, preguntar si el objeto está atornillado a la pared no es una buena idea, porque la respuesta “sí” se obtendrá solo si es el tablero, de modo que su probabilidad es $\frac{1}{8}$ frente a $\frac{7}{8}$ para el “no”. En estos casos lo más adecuado es diseñar un árbol binario de preguntas como el de la figura 10. Dado que hay 8 objetos, se requieren $\log_2(8) = 3$ bits, es decir, 3 niveles de preguntas con respuesta binaria equiprobable, para codificarlos.

En Internet hay juegos que te proponen pensar en un objeto cualquiera de este mundo (20Qnet, 2017), y el *software* lo adivina haciéndote 30 o menos preguntas. Eso significa que el *software* dispone de un árbol binario con 2^{30} , es decir, aproximadamente mil millones de objetos. Es bastante razonable. Con seguridad que lo agotador fue colocar todos estos objetos en el árbol binario del *software*.

La definición de bit de Shannon es la más sencilla que cumple con cuatro propiedades interesantes, que se pueden encontrar en DeDeo (2018): continuidad, simetría, condición de máxima entropía y condición de grano

grueso. Pero no es la única. Hay más de cien definiciones alternativas. Una de las últimas se puede ver en Tsallis (2000).

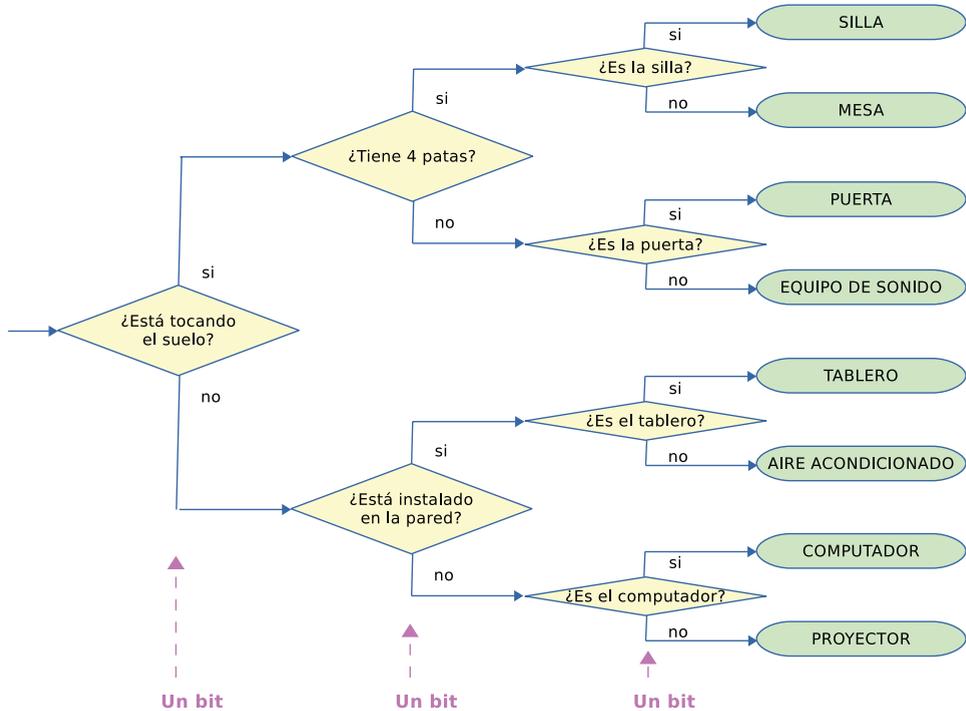


Figura 10. Árbol de preguntas

Las preguntas que debemos hacer para averiguar lo complejo que es un sistema pueden estar orientadas de diversas maneras:

- Medir la **cantidad de información** que describe el objeto haciendo preguntas binarias, como vimos un poco más arriba.
- Medir la **complejidad descriptiva** del objeto, primero escribiendo en español¹³ su descripción. Después se pasa la descripción a una cadena binaria y por último se escribe el programa de computador más corto que imprima esa cadena. La longitud del programa, medida en bits, es la complejidad descriptiva del objeto, y también se llama **complejidad de Kolmogorof-Chaitin** o **información algorítmica**¹⁴ (Chaitin, 1992 y 1997).

13 Para ser exactos, debería usarse un lenguaje formal en vez del español. Pero si las cosas se hacen bien, sin ambigüedades, la diferencia es solo una constante.

14 Dependiendo del libro consultado, estas definiciones de complejidad podrían no ser equivalentes. La idea general es la misma pero podrían diferir en pequeños matices.

- Simular el comportamiento del objeto, usando un programa de computador, y después calcular la **complejidad computacional** (temporal o espacial) de ese programa.
- Medir la **organización interna** del objeto. Se pueden usar medidas termodinámicas (promedios), medidas de caos (el coeficiente de Lyapunov de sensibilidad a errores) o medidas fractales (la dimensión de Hausdorff o el espectro multifractal).
- Medir la **resistencia al cambio** del objeto. Por ejemplo, en un grafo, ¿cuántos arcos escogidos al azar hay que eliminar para que las propiedades del grafo cambien significativamente?
- Hay muchas **otras** medidas de complejidad que no se hacen desde una perspectiva de información, sino combinatoria (Poincaré), como incompletitud (Gödel), como indecidibilidad (Turing), desde la topología (Smale), desde la teoría del caos (Lorenz y Ruelle), como termodinámica del no equilibrio (Prigogine), como geometría fractal (Mandelbrot) o desde la teoría de las catástrofes (Thom).

En todos estos casos hay un problema: no es posible saber si hemos logrado la caracterización más corta posible del objeto. Por ejemplo, si tenemos una descripción de un mecanismo de engranajes que, traducido a binario es:

010001000100010001000100010001000100010001000100010001000100

Quizás el programa más corto que lo imprima sea:

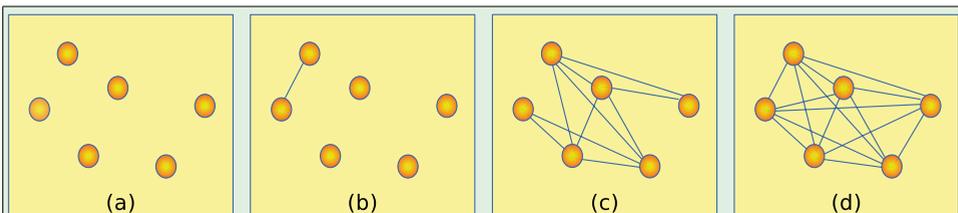
```
print "010001000100010001000100010001000100010001000100010001000100"
```

Que tiene 56 letras (392 bits en código ASCII de 7 bits).

O, si nos damos cuenta que hay regularidades, entonces podríamos escribir:

```
12.times { print "0100" }
```

Que tiene 25 letras, por lo que su complejidad es menor que lo que pensábamos inicialmente (175 bits en ASCII).



Problema 2: ¿Cuál es el sistema más complejo?

Tengo a mi cargo varias salas de computadores (representados por los círculos) y que están interconectados en redes diversas (las conexiones son las líneas).

Atendiendo a las conexiones, ¿cuál es la sala de cómputo más compleja?

¿Habrá algún programa más corto que imprima lo mismo? En Chaitin (1999) se demuestra que no es posible saberlo. Es lo que los matemáticos llaman un indecidible. En vez de poner la demostración formal, lo podemos ver con algunos ejemplos (Problema 3).

Problema 3: Algoritmo descriptor

¿Cuál es la regla de formación de cada una de las siguientes secuencias de números?

- 1, 3, 5, 7, 9, 11, 13, 15, 17, 19
- 1, 4, 1, 5, 9, 2, 6, 5, 3, 5
- 0, 5, 4, 2, 9, 8, 6, 7, 3, 1

Además, la complejidad de un objeto suele depender del nivel de detalle que queremos utilizar. Aunque todos estos objetos representen lo mismo, su nivel de detalle es distinto:

- “Una casa” es una cadena de texto que tiene 8 letras (56 bits de información).



Figura 11. Un icono

- Un icono representando una casa, como el de la figura 11, que ocupa 2.6 KBytes (20.8 Kbits de información).



Figura 12. Casa Batlló, de Gaudí en Barcelona

- Una foto de alta resolución (4043x2988 píxeles con 3 Bytes/píxel para el color), como la de la figura 12, que muestra todos los detalles de una casa sofisticada, y que ocupa 2.9 MBytes (23.2 Mbits de información).

- Una simulación de una casa realizada con un programa de *CAD*, donde se podría entrar y circular por sus espacios, lo cual requeriría muchos más bits de información.
- La casa real, con la posición de cada átomo que la compone.

Una vez definido el nivel de resolución con que vamos a analizar el objeto, hay que tener en cuenta que su complejidad tiene que ver con dos cosas:

- Buscar la forma de comprimir al máximo la información.
- Separar el orden estructural de lo que es puramente aleatorio. Lo aleatorio no se puede comprimir.

Es imposible automatizar estas tareas. La interpretación de los conceptos orden, estructura, complejidad y aleatorio está íntimamente ligada a la inteligencia.



Figura 13. El mismo objeto visto de tres formas, según la inteligencia del observador

En el ejemplo de la figura 13, una abeja ve “muchas cosas verdes” donde un chimpancé ve “un árbol con hojas” y un humano ve una estructura fractal. Cuanto más inteligente se es, más se puede eliminar el ruido, capturar lo estructural y comprimirlo en una descripción lo más corta posible.

Ese es también el objetivo de la ciencia: encontrar fenómenos que se repiten y describirlos con fórmulas matemáticas lo más cortas que se pueda.

La complejidad descriptiva de un objeto no es computable. Es decir, aunque la inmensa mayoría de las cadenas de bits son aleatorias (sin estructura), dada una cadena concreta, no hay forma computable de decidir si es o no aleatoria. Esto fue demostrado por Chaitin, de forma muy parecida a las demostraciones del teorema de indecidibilidad de Gödel y de la parada de Turing. Entonces, cuando calculamos la complejidad de un objeto, estamos manejando una cota superior. El valor real puede ser menor, pero no es computable.

Hay que advertir que, en este contexto, aleatorio y estocástico no es lo mismo:

- Aleatorio: que no se puede comprimir.

- Estocástico: generado por un proceso impredecible. Por ejemplo, una secuencia estocástica de lanzar ocho veces la moneda al aire y que salgan ocho caras (representada por 11111111) es improbable pero posible. Y resulta que esta secuencia concreta sí es comprimible (otras secuencias estocásticas puede que no lo sean).

En el contexto de la computación, emplearíamos los términos pseudoaleatorio y aleatorio, respectivamente, aunque ello vuelve la notación muy confusa. Por ello, a lo largo del libro voy a evitar usar la palabra “aleatorio”, que es ambigua, y solo usaré “pseudoaleatorio” (que no se puede comprimir) o “estocástico” (al azar), que tienen significados precisos y distintos. Además, diferenciaré “pseudoaleatorio” de “caótico” que son muy similares, pero con un pequeño matiz que ya hemos visto en el capítulo sobre caos del libro anterior. Ambos son fenómenos deterministas, es decir, repetibles si se conocen exactamente las condiciones iniciales. Detrás de ellos suele haber una ecuación exacta en la que no interviene el azar. A su vez, “estocástico” y “azar” son sinónimos y representan un tipo de fenómeno que no tiene causas, que no es repetible, que no responde a una ecuación.

Cualquier secuencia de números pseudoaleatoria tiene ciertas garantías matemáticas. Entre las pruebas que debe superar, nunca jamás aparecerá 11111111, pero curiosamente no es en realidad estocástica, sino determinista, aunque usa una fórmula muy compleja.

Si pudiéramos eliminar todo lo estocástico de una cadena de bits que describa a un sistema, lo que queda sería una medida verdadera de su complejidad efectiva. De hecho, la evolución produce esa capacidad de separar lo estocástico de lo repetitivo, de entresacar la información estructural donde aparentemente solo hay ruido.

Podemos ilustrar el problema de diferenciar lo estructural de lo anecdótico (o azaroso) con el siguiente ejemplo: voy en automóvil de mi casa al trabajo y escribo la secuencia de acciones realizadas con los controles del vehículo, siendo:

- Acelerador (A). Donde cada letra A consecutiva significa un 10% más de aceleración. Por ejemplo A significa 0% de aceleración (pedal sin pisar) mientras que AAA significa que el pedal del acelerador está hundido en un 20%.
- Timón (T). El timón puede dirigir el automóvil desde 45 grados hacia la izquierda hasta 45 grados hacia la derecha, en 10 saltos de 9 grados cada uno. De este modo, TTTTTT significa ir de frente, TTTT significa girar ligeramente a la izquierda y TTTTTT significa girar ligeramente a la derecha.

- Palanca de Cambio de marchas (C). Poner la palanca de cambio en primera se codifica C, en segunda CC, en tercera CCC, en cuarta CCCC, en neutro CCCCC y en reversa CCCCCC.
- Grado de frenada (G). Donde cada letra G consecutiva significa un 10% más de frenada. Por ejemplo, G significa 0% de frenada (no estoy pisando el pedal), mientras que GGGGG significa un 50% de frenada.

De este modo, cada vez que realice el recorrido obtendré una secuencia de letras ATCG.

Encontraremos secuencias que se repiten con mucha frecuencia como CCCCCGGGGG que significa que estamos detenidos (seguramente esperando en un semáforo rojo), o GCAACCAAACCCAAAA que significa que estamos arrancando (típicamente cuando el semáforo se pone verde).

Si nos empeñamos en buscarlos, encontraremos muchos patrones, pero todo ello es inútil por lo anecdótico. Estamos analizando un sistema a un nivel que tiene demasiados detalles descriptivos al azar, que realmente no indican nada. Todos los días logro mi objetivo de ir de casa al trabajo, pero cada día ocurren cosas imprevisibles que alteran mis acciones sobre el vehículo (un atasco que me obliga a cambiar de ruta, un gato que se atraviesa y tengo que frenar bruscamente, tráfico lento que me obliga a ir más despacio, etc.). De ese conjunto de secuencias no podremos deducir ninguna información (que me lleva de la casa al trabajo) hasta que logremos eliminar todo ese ruido.

LA GRADUALIDAD Y LA EMERGENCIA

Si bien estamos argumentando que la vida, la inteligencia, la consciencia y la libertad no son cuestiones de todo o nada sino graduales, tampoco tienen por qué crecer linealmente. Puede haber saltos. Los saltos se originan habitualmente por realimentaciones positivas y negativas (Figura 14).

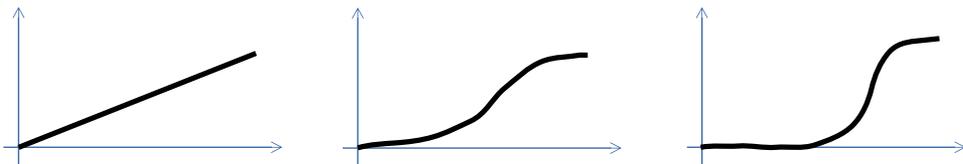


Figura 14. Crecimiento de la complejidad: a) lineal; b) no lineal; c) francamente exponencial con saturación.

Hay zonas de crecimiento aparentemente lineal, hasta que se llega al punto de desbordamiento, o de emergencia en las cuales el crecimiento es expo-

nencial, hasta que se agota algún recurso y el sistema se vuelve a estabilizar, pero allí aparece un nuevo bucle y poco a poco va aumentando otra vez la complejidad (Figura 15). Y ello puede ocurrir varias veces. De hecho, las zonas de aparente estancamiento o de crecimiento lineal muy bien pueden ser el comienzo de la exponencial, que al principio es muy lenta. Al respecto, el crecimiento exponencial de la tecnología que estamos viviendo comenzó en el Neolítico, cuando el *Homo Sapiens* comenzó a fabricar las primeras herramientas, pero seguramente no se daba cuenta de ello: pasaron miles de años y la vida seguía más o menos igual. Hasta la época de nuestros tatarabuelos el crecimiento podría parecer cero o a lo mucho lineal, pues ellos apenas tuvieron que aprender algo nuevo en su vida, como la radio. No llegó ningún aparato más a la casa en toda su vida, por contraste con nuestros tiempos, que llegan nuevos *gadgets* cada pocos meses.

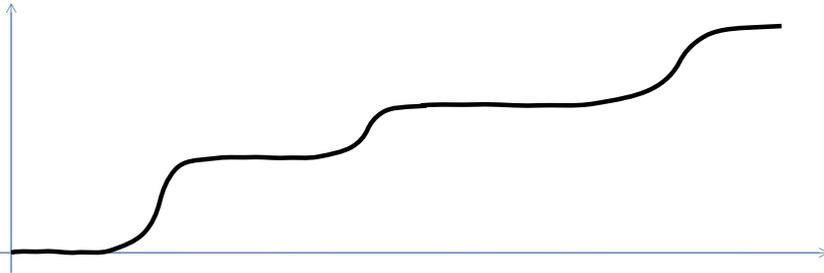


Figura 15. Sucesivos crecimientos exponenciales de la complejidad

La emergencia es un fenómeno exponencial. Se acumulan interacciones en el nivel inferior, que producen una avalancha de más interacciones, gracias a algún bucle de realimentación positiva. Al final, aparece un bucle de realimentación negativa (puede ser una simple saturación) que estabiliza el sistema. Pero ha dejado una huella, una acumulación de novedades que crea una perspectiva distinta de ver las cosas. Aunque hay que matizar esto: la mayoría de las veces la emergencia es solo una cuestión subjetiva de interpretación, que crea una nueva ontología de objetos y relaciones, que antes no existía, pero solo a los ojos de un observador inteligente. Por ejemplo, cuando tenemos muchas moléculas gaseosas de oxígeno, nitrógeno, agua y anhídrido carbónico, ellas chocan entre sí y rebotan, siguiendo las leyes de la física, de la conservación del momento. Cuando hay una cantidad enorme de estas moléculas, emergen unas propiedades medibles por nosotros, como volumen, presión y temperatura, pero las moléculas siguen interactuando entre sí según la ley de conservación del momento. No comienzan a obedecer nuevas leyes de alto nivel relativas a temperatura y presión. Nosotros si

podemos comenzar a explicar y predecir los sistemas con los nuevos conceptos. Por ejemplo, si mezclamos dos recipientes con aire, cada uno a cierta temperatura, podemos saber a qué temperatura quedará el aire resultante sin necesidad de calcular los momentos de cada molécula. Hay una nueva ontología y unas nuevas reglas de interacción. Pero son solo aproximadas, pueden ser incluso una excelente aproximación, pero las reglas que operan realmente siguen siendo las de las colisiones entre moléculas que conservan su momento.

Y si surge alguna discrepancia entre las reglas de alto y bajo nivel, son las de bajo nivel las que realmente se cumplen. Quizás por eso en alto nivel aparecen fenómenos aparentemente estocásticos, cuando sus predicciones se apartan de lo que realmente ocurre en el bajo nivel.

Sin embargo, también hay veces donde la emergencia deja de ser una mera cuestión de interpretación. Y eso ocurre cuando aparecen nuevos objetos con relaciones nuevas entre ellos. Más adelante veremos un ejemplo con el concepto del dinero.

Regresando al tema de la gradualidad, cuando una propiedad crece linealmente, y de la emergencia, cuando crece exponencialmente, si decimos que un destornillador es inteligente (Figura 2), no queremos decir que sepa detectar sus errores, corregirlos, aprender nuevas técnicas o incluso adaptarse a resolver nuevos problemas. Todo ello es producto de emergencias de muy alto nivel. El destornillador es inteligente porque sirve para algo: para fijar tornillos. En ese sentido, una piedra es menos inteligente que un destornillador. Pero en cualquier caso estamos muy lejos de la inteligencia en el sentido humano, que es también un fenómeno emergente.

Como resumen de lo que es la emergencia, podemos decir que hay un crecimiento gradual de nuevas propiedades a la vez que, de vez en cuando, hay emergencia exponencial de otras nuevas. Esto ha sido motivo de discusiones apasionadas en teoría de la evolución entre los gradualistas (como el propio Darwin) frente a los puntualistas (como Jay Gould) que no tienen mucho sentido, pues ambos tienen razón. Es habitual que se den ambas cosas simultáneamente y en muchos niveles. Debido a la competencia, poco a poco se van refinando y mejorando características ya existentes. Eso es lo que se conoce como evolución gradual, que ocurre a lo largo de cientos, miles e incluso millones de años, cuando es de tipo biológico. La competencia es un algoritmo optimizador (como hemos visto en el capítulo de algoritmos evolutivos del libro anterior) y produce mejoras graduales. Por el contrario, la cooperación es emergente: cuando ocurren por accidente situaciones donde dos o más individuos se ven obligados a cooperar para mejor sobrevivir y reproducirse, entonces aparecen nuevas estructuras que antes

no existían. En biología, la cooperación es lo que logra los grandes saltos evolutivos, como el paso de las células procariotas a las eucariotas, muy bien explicado por Lynn Margulis en el libro *¿Qué es el sexo?* Y si quieres conocer todos los saltos evolutivos que se han dado hasta ahora, te recomiendo leer el libro *Ocho hitos de la evolución*, de John Maynard Smith.

¿LA COMPLEJIDAD PUEDE CRECER INDEFINIDAMENTE?

No parece que sea este el caso aun cuando no hay ningún límite prefijado de antemano. En el 2001 con José Alejandro Gómez, en EVALAB, realizamos una serie de experimentos con algoritmos coevolutivos, donde hay muchos agentes cooperando o compitiendo entre sí. Y vimos que la cooperación se construía, y podía aumentar y durar mucho tiempo, pero sin ninguna razón aparente llegaba un momento en que se destruía por completo, y el proceso volvía a comenzar. La forma de hacerlo recordaba las extinciones masivas que ha sufrido la vida en nuestro planeta.

En sistemas sociales parece ocurrir lo mismo. Muchas civilizaciones se extinguieron sin ninguna razón aparente, como la civilización Maya en México y Centroamérica, que colapsó en el siglo IX en un momento de gran prosperidad. Las gentes abandonaron las ciudades sin que se sepa la causa real, aunque probablemente afrontaron problemas de superpoblación y sequías, teniendo unos gobernantes que daban respuestas tradicionales a problemas que eran nuevos. Algo similar ocurrió con la civilización del Valle del Indo en la Edad del Bronce, que también desapareció abruptamente. La caída del imperio romano está más documentada pero aún hay mucha discusión sobre las causas, que parecen más internas (la corrupción, enfermedades, hambre y un imperio demasiado grande para la velocidad de comunicaciones de la época) que externas (los bárbaros).

Roger Lewin en su libro *Complexity* documenta muchos de estos colapsos, principalmente de culturas norteamericanas como la del Chaco. Lewin nos hace notar que suele haber un patrón común: las sociedades incrementan sus actividades, principalmente la construcción, justo antes de colapsar. En estas civilizaciones, la etapa previa al Estado es muy estable y cuando finalmente surge el Estado, viene precedido de un colapso. Es decir, no necesariamente el colapso lleva a la desaparición de la sociedad, sino que puede conducir a otra forma de organización social.

William Irwin Thompson (2001) insiste más en el mismo fenómeno: es común en la historia ver cómo las transformaciones radicales son precedidas por una intensificación de las posiciones antiguas. Por ejemplo, en la Edad Media cuando la armadura se hace tan sofisticada que cada parte del cuerpo

está protegida con algún accesorio, y a causa de ello es tan pesada que para elevar a los caballeros sobre los caballos se emplean mecanismos de palancas y poleas; pues bien, es justo en ese momento en que la armadura se vuelve obsoleta, debido a la invención de las armas de fuego, desapareciendo el feudalismo y dando paso a los Estados nación. Thompson llama a esto el efecto de ocaso o de supernova. Y lo más probable es que se trate no tanto de una intensificación, sino de una exponencial volviéndose cada vez más rápida. En la figura 16 representamos un crecimiento de complejidad exponencial que justo antes de explotar y destruir el sistema (la línea a trazos) consigue convertirse en otra cosa, que a su vez sufre otro crecimiento exponencial, y así sucesivamente, pero cada vez más deprisa. Es muy similar a la figura 15, pero en esta las exponenciales terminan en saturación, mientras que en aquella terminan en destrucción. Esta figura también representa muy bien lo que ocurre con las actuales tecnologías electrónicas, informáticas y de comunicaciones. Por ejemplo, de la escritura en piedra a la imprenta, a la máquina de escribir, a los teletipos, a los computadores, a los portátiles, tabletas y teléfonos móviles... Las transiciones son cada vez más rápidas. Todas las empresas fabricantes deben innovar cada vez más deprisa. Recuerdo cuando en 1970 se comenzó a hablar de las memorias de burbujas magnéticas y que IBM dijo que esperarían 10 años para instalarlas en sus computadores porque no veían en ese momento necesidad de mayores capacidades. Suena gracioso porque si hoy se te ocurre una gran idea, hay que sacarla a producción en menos de seis meses si no quieres que la competencia te adelante.

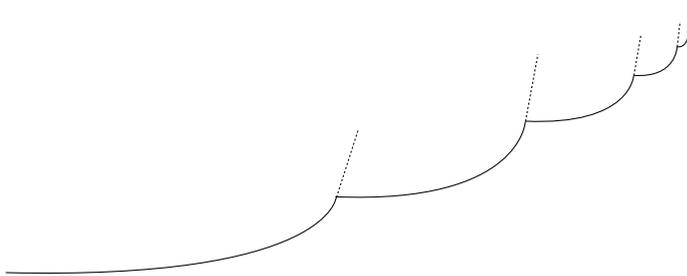


Figura 16. Sucesivos crecimientos exponenciales justo antes de la catástrofe

Y en esta figura donde se evita el final brusco de las exponenciales conmutando a otras exponenciales cada vez más rápidas, se vislumbra un final nada halagüeño.

De hecho, hay varios visionarios que ya se han dado cuenta de que nos espera un problema gordo alrededor del año 2050: lo llaman “la gran singularidad” y tiene que ver con que para esa fecha aproximada un computador,

siguiendo la ley empírica de Moore¹⁵, habrá alcanzado la capacidad de cómputo equivalente a todos los cerebros humanos juntos. Y eso tiene que tener consecuencias importantes para el funcionamiento de las sociedades, aun cuando ahora mismo no podamos imaginar cuáles van a ser.

No obstante, se esperan más singularidades alrededor de esa fecha, como que se acaben el petróleo y el agua potable. Los principales divulgadores de esta idea son Ray Kurzweil y Peter Diamandis, que fundaron en *Silicon Valley* la *Singularity University* con ayuda financiera de Google y NASA. En matemáticas “singularidad” significa “infinito” y esa es la idea: hay algo que está creciendo o decreciendo de forma exponencial para terminar en una singularidad. ¿Qué va a pasar entonces? Probablemente un colapso de la civilización con todo tipo de guerras, que quizás dé lugar a otro tipo de organización social, como señalan Lewin y Thompson. Pero Kurzweil y Diamandis son optimistas y piensan que nos espera un futuro completamente tecnológico, con todas las necesidades de supervivencia satisfechas y donde nuestra imaginación es el límite. Un futuro de abundancia. ¿Vamos a esperar a ver quién tiene razón? ¿No será más inteligente anticiparnos y construir la solución más razonable para la humanidad? Kurzweil actualmente dirige una división de Inteligencia Artificial de Google, y piensa que el futuro está ahí: las máquinas realizarán todo el trabajo, incluso probablemente serán las administradoras de la sociedad (quizás eliminando así la corrupción en la política). Obviamente la gente no querrá morir de hambre porque se quede sin trabajo, y los empresarios no querrán ver quebrar sus empresas porque ya nadie tiene poder adquisitivo para comprar los bienes que producen. La solución a este problema puede ser el salario universal, un salario razonable que el Estado da a todas las personas tanto si trabajan como si no. Ya se han hecho experimentos al respecto, algunos con mucho éxito, en ciertas regiones de Canadá, en Uganda, en Brasil, Finlandia, Alaska, Escocia y Holanda. Hace unos meses, sorprendentemente, Suiza lo rechazó en un referéndum. Mucha gente cree que no es necesario prepararse para ese futuro, que los computadores no llegarán a desplazar a los humanos porque la creatividad humana no se puede imitar, que las cosas siempre seguirán igual. Todo eso es falso. Precisamente los algoritmos evolutivos son fuertemente creativos y han conseguido resultados que emulan y superan a los de los humanos. Si la inercia social nos impide reaccionar a tiempo, eso sí nos puede llevar a la catástrofe.

15 Que dice que el número de transistores en un chip se duplica cada 2 años, a la vez que el chip se hace más rápido y barato. Es decir, la capacidad de cómputo del silicio crece exponencialmente.

Y ¿cuáles son los mecanismos que llevan los sistemas complejos a su destrucción? Se han identificado tres formas, que son meros aspectos de lo mismo:

- Para aumentar su complejidad usan muchos recursos, socavando los del nivel anterior. Por ejemplo, en la Isla de Pascua se supone que los habitantes se enzarzaron en competencias para fabricar más y más estatuas *moái* de modo que acabaron con los árboles que usaban para transportarlas y sin los cuales no pudieron construir más canoas para pescar. Otro ejemplo en los tiempos actuales, es que para poder construir más y más objetos industriales, que se venden muy bien, estamos envenenando el aire y el agua.
- La complejidad es tan alta que la gente ya no puede entenderla (es lo mismo que el punto anterior, con la salvedad de que los recursos que se acaban no son físicos sino intelectuales). Se produce entonces un auge de los populismos y de las ideas esotéricas porque la gente prefiere mentiras sencillas a verdades complejas. Con ello la sociedad compleja y sofisticada se resquebraja en trozos y se destruye. Posiblemente eso es lo que está ocurriendo ahora mismo con el Brexit, la elección de Trump y el resurgimiento de nacionalismos. En lo personal, puedo mencionar como anécdota que, hablando con gente que vive a la orilla del mar, muchos no creen en el cambio climático con el consiguiente derretimiento del continente Antártico, porque en toda su vida el nivel del agua no ha aumentado. Sin embargo, y aunque ellos no son conscientes, en los últimos 100 años ya ha subido unos 15 cm. El problema de hacer la medida a ojo es que el mar no es plano. Tiene olas, corrientes, movimientos por vientos y mareas, tormentas. Además, la subida es muy lenta para apreciarla en unos pocos años. Una persona no puede medir si ha subido o no, ni cuánto. La ciencia está usando diversos métodos y ahora tenemos uno preciso que son las mediciones desde satélites. Y ellas indican una subida actual de 3 mm/año que, además, no es lineal como se creía al principio, sino acelerada, como corresponde a cualquier fenómeno complejo con realimentación positiva. Todo ello es muy difícil de entender para una persona común. Es más simple suponer que sus ojos no le engañan y que todo sigue igual. Las mentiras simples triunfan sobre las verdades complejas.
- El aumento de la complejidad, generando estructuras que dependen fuertemente de las antiguas, vuelve el sistema muy frágil. Cualquier incidente externo puede destruir el sistema. Esto pasa con el oso panda de China, que se especializó tanto en comer bambú, que puede extinguirse si desaparece esta planta. Eso puede pasar con nuestros

sistemas eléctricos, electrónicos y de comunicaciones en los que se basa por completo nuestra actual sociedad de la información. Una tormenta solar fuerte (o una bomba atómica electromagnética) podría quemar la mayoría de los equipos, incluyendo sus fábricas, y devolvernos al Neolítico. Recordemos que un fallo de suministro eléctrico en 1995 en una fábrica de semiconductores en una isla de Malasia o un simple incendio en otra fábrica en 2013 en China afectaron fuertemente a la industria mundial de computadores.

Al principio la evolución genera complejidad creciente pero tampoco está garantizado que perdure. Es bien sabido entre los ingenieros que usamos algoritmos evolutivos, que funcionan mal al final, cuando se aproximan a una buena solución, porque entonces cualquier mejora que hagan es pequeña y la presión selectiva, que se basa precisamente en eso, también es pequeña. Especialmente si se han formado sociedades con cooperación interna, la población puede crecer exponencialmente sin ningún freno. Entonces, cuando las poblaciones son grandes, puede haber pequeños grupos de individuos que se muevan al azar o incluso en dirección contraria a la presión selectiva, sin que por ello vayan a desaparecer, ya que el resto de la comunidad los mantiene. Eso incluye cosas tan diversas, pero todas contrarias a la evolución, como el celibato, el control de la natalidad decidido por el individuo o por el Estado, el suicidio colectivo o el terrorismo. La razón de que estos fenómenos prosperen es el tiempo libre, la superpoblación y las temporizaciones. Veamos lo que significa cada factor.

Tiempo libre. Cuando superas de forma sistemática y automática los retos que la vida te pone por delante, entonces te queda tiempo libre. El tiempo libre permite hacer actividades raras, no relacionadas con la supervivencia ni la reproducción, como por ejemplo pintar en las paredes de las cuevas y construir naves espaciales. En muchos casos, esos subproductos se enredan culturalmente con los genes para apoyar su objetivo de supervivencia. Es así como recitar un poema puede ser el prelude de una boda, o desarrollar tecnologías médicas puede alargar vidas. Y quizás las naves espaciales puedan evitar una extinción masiva. Pero en otros casos esos subproductos pueden ir en contra de la supervivencia de los genes, y así hemos creado artefactos de guerra encaminados a exterminar toda la vida del planeta.

Superpoblación. Cuando tienes superpoblación en un ambiente social, si hay un grupo pequeño de individuos que no se reproducen o incluso que se matan entre sí, ello no se nota a corto plazo pues el grueso de la población seguirá allí y, por tanto, no habrá ninguna presión evolutiva para evitarlo. Lo malo es que las ideas del grupo pueden propagarse con mucha rapidez si

se da el ambiente adecuado, como ocurre con los populismos políticos, de manera que cuando comiencen a ser evidentes sus efectos será demasiado tarde para frenarlas, pudiendo destruir por completo la sociedad. La superpoblación enmascara muchos problemas.

Temporizaciones. Los retrasos temporales también juegan un papel importante: la constante temporal de la evolución es el tiempo entre generaciones. Por ejemplo, en los humanos está entre 20 y 30 años, mientras que en las bacterias es cuestión de unas pocas horas. Ese es de alguna manera el periodo de muestreo donde el algoritmo evolutivo ejerce su control. Pero puede haber fenómenos más lentos o más rápidos que escapen a ese control:

- Más lentos: cuando se inicia un proceso que no tiene importancia en las generaciones inmediatas, pero que traerá consecuencias en las generaciones del futuro lejano. Por ejemplo, la contaminación del planeta: la evolución no puede evitarla porque es un proceso exponencial que comienza de forma imperceptible, de modo que no impide tener hijos y nietos a los individuos actuales. No obstante, afectará adversamente a los tataranietos, cuando la exponencial crezca y no haya forma de pararla.
- Más rápidos: una persona nace, crea arte, tiene hijos y muere. La evolución no favorece ni impide el arte, que se hace a una velocidad más rápida. El algoritmo evolutivo no nota que estamos empleando el tiempo libre en cosas que no le interesan pues no afectan a la supervivencia ni a tener hijos. Sin embargo, en el tiempo libre puedes hacer cosas que interfieran a largo plazo con la evolución, como fabricar bombas atómicas.

Otro ejemplo: la ciencia es un sistema evolutivo, ya que hay una colección de ideas que se reproducen haciendo pequeños cambios o combinándose con otras y sometidas a la presión selectiva de la evaluación de pares para alcanzar su publicación. Sin embargo, se ha constatado que el fraude en ciencia va en aumento (Smaldino y McElreath, 2017). Las razones se acaban de explicar: hay superpoblación de artículos en revistas y congresos, de modo que inicialmente el fraude no supone un grave perjuicio para todo el sistema, que sigue funcionando. Aunque con el paso del tiempo puede colapsar, ya que las investigaciones se hacen confiando en la verdad de los resultados publicados anteriormente. Pero ese proceso es mucho más lento que el evolutivo. Mientras tanto hay unos beneficios inmediatos para el investigador, quien recibe recompensas salariales y de prestigio cuanto más publique. Este otro proceso es mucho más rápido que el evolutivo. Y, como consecuencia, la ciencia puede colapsar.

EMERGENCIA, TRANSICIONES DE FASE Y AUTOORGANIZACIÓN

Son conceptos muy similares. Se dice que ha habido emergencia cuando de la interacción de individuos aparece una nueva propiedad que no estaba en ninguno de ellos por separado. Mientras que cuando hablamos de transiciones de fase¹⁶ estamos explicando el mismo fenómeno pero haciendo énfasis en que hay un umbral en la cantidad de individuos que produce el cambio. Por debajo del umbral la población se encuentra en un estado (o fase) y por encima, en otro.

Cuando se habla de autoorganización es porque un conjunto de individuos se reúne y colabora para formar un superorganismo que preserva su propia frontera, lo cual también es un fenómeno emergente y también requiere superar un umbral, pero ahora hacemos énfasis en que aparece un superorganismo que antes no existía. Por ejemplo, un conjunto de moléculas forma una célula; un conjunto de células forma un animal; un conjunto de pájaros forma una bandada. Una característica importante de este fenómeno es que no hay un control central que decida cómo van a interactuar los entes. Todo ocurre de forma distribuida, redundante y con tolerancia a fallos (no le pasa nada a la bandada si un pájaro cambia de lugar o se muere). La autoorganización es productora de neguentropía, o entropía negativa¹⁷.

Y aunque hay muchos ejemplos de sistemas autoorganizados, tanto biológicos como computacionales, no se ha hecho un estudio acerca de cuántos bits de complejidad se requieren para que aparezca este fenómeno. Pero no parecen ser muchos porque lo único que se requiere son unas simples reglas de interacción entre individuos.



Figura 17. Autoorganización de una bandada de pájaros en el cielo de Madrid

16 Término tomado de la Física, por analogía a las transiciones de fase entre líquido y sólido.

17 Claro, a nivel local, pues ya sabemos que globalmente la entropía siempre aumenta.

Desde luego, si ya existe la evolución, lograr agrupación de entes (o lograr cualquier otra cosa) es relativamente sencillo. En el video del profesor William C. Ratcliff del 2014 que se encuentra en las referencias, podemos ver el experimento de cómo pasar de células aisladas a un cuerpo multicelular sencillo. En general, si tienes una población que se reproduce y hay pequeños errores en las copias, puedes guiar la evolución a que cumpla casi que cualquier objetivo, manipulando adecuadamente la función de aptitud, es decir, decidiendo quiénes son los que se reproducen. Pero incluso si no usamos ningún algoritmo evolutivo, la autoorganización puede surgir de la combinación de reglas simples. Y eso es lo que vemos cuando los pájaros forman bandadas (Figura 17): un observador ingenuo podría pensar que hay un pájaro líder que da las órdenes de cómo deben seguirlo los demás pájaros, pero los biólogos saben que eso no es así. No hay ningún líder en las grandes bandadas de pájaros. Hace ya mucho, en 1987, Craig Reynolds escribió *Boids*, un modelo de *software* en el que se puede observar el mismo fenómeno si cada pájaro sigue básicamente dos reglas:

- Moverse en la dirección y velocidad promedio de los pájaros que tiene a su alrededor.
- Evitar choques con otros pájaros u obstáculos.

Entonces, de la interacción de esas dos simples reglas operando en un gran número de individuos surge un fenómeno emergente de autoorganización: la bandada.

Los cardúmenes de peces se forman usando unas reglas parecidas, a las que se añade otra: moverse hacia el centro, con lo cual evitan estar en el borde exterior del cardumen, que es donde se corre más peligro de ser devorado por peces mayores. El resultado de ello es un movimiento característico del cardumen, en rotación continua hacia adentro.

Otro ejemplo de emergencia en poblaciones de animales lo encontramos en Miller (2010). La langosta del desierto del norte de África es caníbal y tiene dos modos de funcionamiento: cuando hay pocas, se mantienen alejadas entre sí comiendo vegetales. Mientras que cuando llueve y las larvas de los huevos encuentran comida de sobra se produce superpoblación. Entonces se encuentran demasiado juntas pero de forma inestable, pues cada una trata de comerse a la que tiene delante, a la vez que tratan de evadir a las que están detrás. Por ello el miedo las empuja a saltar continuamente hacia adelante y allí donde caen producen pánico que a su vez obliga a saltar a las siguientes, en un fenómeno de realimentación positiva. Ahí es donde se convierten una plaga temible, que acaba con todo lo que encuentran. Es interesante saber que la hormona que regula esta transición de fase en la langosta es la sero-

tonina, que también la tenemos los humanos, aunque todavía no se sabe si puede producir efectos similares.

Además, no solo hay autoorganización en sistemas vivos. También existe en cualquier otro sistema donde haya muchas partes que interactúen entre sí de forma no lineal. Los ejemplos clásicos son las dunas (Figura 18), que se forman por la interacción del viento con muchas partículas de arena, y que producen pendientes y formas muy similares a pesar de que no hay comunicación entre ellas. Las dunas son un ejemplo de criticalidad autoorganizada que se produce cuando hay al superarse algún umbral aparece una realimentación positiva. La pendiente de las dunas es de un 35% aproximadamente. Cuando alcanza valores mayores se derrumba porque los granos de arena no pueden soportarse unos sobre otros, y en su caída pueden golpear y arrastrar a otros. Esa es la realimentación positiva. Debido a consideraciones físicas y geométricas emerge esa pendiente que suele ser algo menor a 35% para que sea estable, es decir, para que no ocurra la realimentación positiva. Además, las avalanchas que aparecen siguen una ley de potencias con exponente cercano a 2, dependiendo del modelo y condiciones usadas. Hay muchas avalanchas pequeñas y pocas grandes.



Figura 18. Dunas y microdunas en Mangue Seco, Brasil

Otro ejemplo de criticalidad es el contenido de oxígeno de la atmósfera, que es del 21%. Si alcanzase el 25% se producirían incendios imposibles de apagar, pero entonces el oxígeno se consumiría hasta alcanzar un nivel seguro. Por ello lo habitual es ver niveles un poco menores a 25%. Otro ejemplo descubierto muy recientemente es la masa del bosón de Higgs, que es de 126Gev,

lo cual produce un universo al borde del colapso. Si se superase ese umbral no observaríamos ningún universo ya que ni siquiera existiríamos nosotros.

Alan Turing también realizó trabajos valiosos sobre la emergencia de patrones en la piel de los animales, que no vieron la luz en su momento sino solo hace unos pocos años pues, al parecer, no llamaron la atención de su director de tesis doctoral. Las cebras y los tigres tienen rayas, mientras que el guepardo tiene manchas y muchos peces tropicales, ranas y serpientes también exhiben patrones coloreados. Debería ser obvio que el cerebro de los animales no se encarga de guiar esta coloración pues tiene cosas más importantes en que ocuparse, de modo que el cómputo es local: una célula de la piel decide ponerse de un color o de otro en función de lo que hagan sus vecinas. Turing propuso un modelo de reacción-difusión (resumido en la ecuación 4), que es continuo y que trata de capturar las reacciones químicas que ocurren cuando se encuentran dos o más sustancias, así como su difusión cuando hay diferentes concentraciones. Este modelo da cuenta de muchos patrones ordenados que surgen en la naturaleza. Según lo explica Peña (2002), consiste en dos sustancias que reaccionan entre sí, pero que también se difunden, es decir, se mueven de los lugares donde hay más concentración hacia donde hay menos. Suponemos que una de las sustancias es autocatalítica, es decir, que genera la producción de más de sí misma, o sea, se trata de un bucle de realimentación positiva. Esta sustancia también produce por reacciones una segunda sustancia, que es inhibidora de la primera, conformando un bucle de realimentación negativa. Ambas sustancias se difunden espacialmente pero a velocidades distintas y si en una zona hay más concentración de una sustancia que de otra, la piel del animal adquiere un color u otro. Al respecto, quizás recordemos del libro anterior el ejemplo de autómatas celulares llamado “bandera francesa”, donde era crítica la difusión de señales de colores a distinta velocidad, junto con un inhibidor.

$$\frac{\partial C}{\partial t} = D\nabla^2 C + f(C) \quad \text{Ec. 4}$$

En este modelo, C es el vector de concentraciones de sustancias, D es la matriz diagonal con los coeficientes de difusión de cada una y $f()$ es un vector que especifica las reacciones químicas entre las dos sustancias, es decir, son ecuaciones no lineales que indican cuanto se produce de cada sustancia en función de las concentraciones actuales de cada una.

James Dickson Murray fue quien desarrolló el trabajo más exhaustivo al resolver estas ecuaciones de forma aproximada, en función de los valores que se asignen a las constantes y aplicado a varios escenarios biológicos. Al hacerlo con la piel de los animales se producen soluciones donde el co-

lor es homogéneo, con manchas (vacas, guepardos) o rayas (cebras, tigres). Ciertos parámetros controlan lo grande o pequeño de las manchas o rayas. Al hacer el análisis matemático del problema, Murray (2002, vol. II, p. 147) enunció un teorema famoso por lo raro: la cola de los animales no puede terminar en manchas, aunque sí en rayas, debido a que son las únicas soluciones estables en espacios unidimensionales (las colas son aproximadamente unidimensionales).



Figura 19. Patrón rotacional en (a) aguacate; (b) madreperla; (c) bromelia; (d) maleza. Las hojas están numeradas, siendo 1 la más joven

El mismo modelo sirve para entender cómo crece un embrión a partir de sus células germinales iniciales. Predice que las extremidades de los animales terminan en cinco dedos, lo cual se ve incluso en las articulaciones de las aletas de las ballenas (Goodwin, 1998).

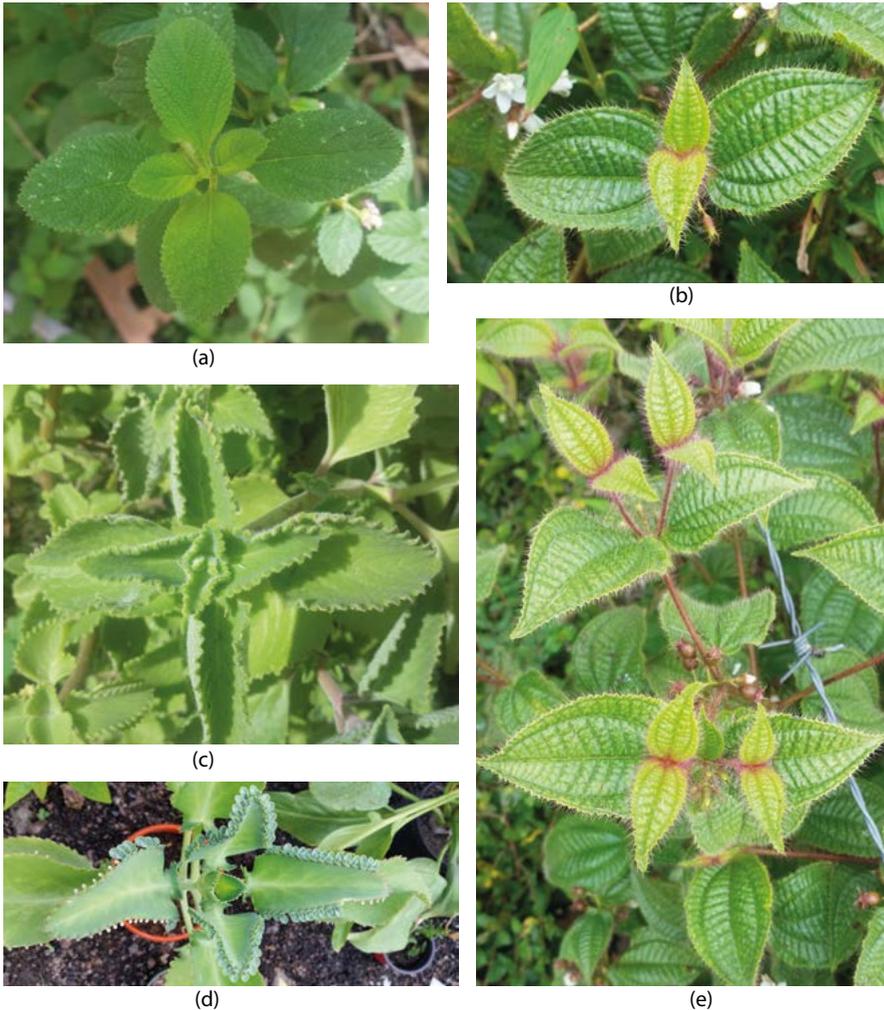


Figura 20. Patrón de oposición en (a) prontoalivio; (b) pipilonga (c); oreganón; (d) lengua de la suegra; (e) otro tipo de pipilonga

También explican el crecimiento de las hojas de muchas plantas que pueden aparecer según unos pocos patrones como el rotacional (Figura 19) y el de oposición (Figura 20), ambas muy bien explicadas por Brian Goodwin (1998) en su libro *Las manchas del leopardo*.

Aunque los biólogos no se ponen de acuerdo en los detalles, parecería que ambas estrategias corresponden a un proceso para maximizar la luz recibida. Las hojas recién nacidas son más pequeñas pero, en cualquier caso, proyectan sombra sobre las hojas más antiguas. Para el patrón rotacional el ángulo de giro es de esperar que sea un número irracional, y así ninguna hoja quedará exactamente encima de ninguna otra. Efectivamente en mu-

chas plantas sale el ángulo áureo, que es irracional trascendente, y que equivale al número áureo pero expresado en grados (aproximadamente 137.5 grados). La figura 21 y la ecuación 5 nos ayudan a recordar de donde salen estos números. Pero no hay garantías de que todas las plantas hagan lo mismo, como podemos ver en la figura 19-d, donde el patrón se repite pero el ángulo no es el áureo, y ello significa que los genes pueden haber encontrado otras soluciones en función de circunstancias difíciles de imaginar.



Figura 21. Número áureo y ángulo áureo

El número áureo tiene una gran ventaja y seguramente por eso es que aparece frecuentemente en la naturaleza: permite usar el mismo algoritmo para generar estructuras autosimilares a distintas escalas, donde permanezca la misma relación entre dos magnitudes. Como vemos en la ecuación 5, en un objeto con dos partes a y b , la relación del todo ($a+b$) respecto a la parte mayor (a) es la misma que la relación entre la parte mayor (a) respecto a la parte más pequeña (b). Son algoritmos recursivos, que producen fractales.

$$\frac{a}{b} = \frac{a+b}{a} \Rightarrow \text{phi} = \frac{a}{b} = 1.6180339 \dots$$

$$\frac{\alpha}{\beta} = \frac{\alpha+\beta}{\alpha} \quad \text{con } \alpha+\beta=360^\circ \Rightarrow \beta=137.507 \dots^\circ$$

Ec. 5

Si efectivamente es un proceso de optimización, quienes lo dirigen son los genes. No hay otra alternativa. Pero no tienen mucha libertad. Están restringidos a un espacio de formas matemático, y también encontrarán otras restricciones físicas y químicas. Un ejemplo de restricción física es que las hojas no pueden ser tan grandes que los tallos se rompan por su peso. Un ejemplo de restricción físico-químico-biológico es la imposibilidad de que las hojas floten libremente en el aire y transmitan a la planta la energía solar recibida vía enlaces de microondas, como planeamos hacer los humanos con paneles solares en órbita. Los genes no saben cómo hacer algo tan complejo, aun cuando pudiera ser óptimo. De la misma manera, los genes inventaron las patas de los animales para caminar, pero no las ruedas, porque es

imposible mantener vivo una parte del cuerpo si no se le pueden transmitir fluidos como la sangre o la savia, si está girando continuamente, de modo que cualquier tubo de transporte se retorcería y estrangularía al instante.

El patrón de oposición es más sencillo de entender y de ver, pues las hojas nacen por pares que se dirigen en direcciones opuestas. El siguiente par hace lo mismo, aunque está girado 90 grados respecto al par anterior. También así se logra recibir más luz solar. De nuevo, ante un patrón matemático evidente los genes pueden hacer variaciones, como vemos en la figura 20-e que nacen simultáneamente dos pares de hojas, en vez de uno solo. Y también se generan otras formas que podrían interpretarse como inicio de fractales, como en la figura 20-d que hay hojas en los bordes de las hojas. Los fractales vegetales son un tema extenso, abordado con mucha profundidad por Prusinkiewicz y Lindenmayer (2004) en su famoso libro *La belleza algorítmica de las plantas*.

Por ello, a esta propuesta de Turing se le llama modelo de morfogénesis, es decir, de generación de formas, lo cual —por cierto— dio lugar a un corto debate sobre si la evolución la controlaban los genes o las matemáticas de esta ecuación. La respuesta es que son las matemáticas las que generan las formas iniciales, mientras que los genes se limitan a introducir variaciones sobre el tema principal. Variaciones que al final pueden ser bastante grandes. Resaltemos que los genes están presentes únicamente en cuerpos biológicos, mientras que las ecuaciones de reacción-difusión actúan también en estructuras previas a la vida, como reacciones químicas y procesos geológicos que vamos a ver enseguida.

Beloúsov y Zhabotinski encontraron patrones en ciertas reacciones químicas que se creían imposibles en su época (Figura 22), por lo que sus trabajos fueron rechazados en las revistas durante mucho tiempo. Se trata de reacciones químicas reversibles, también llamadas reacciones-reloj, que son caóticas en el sentido débil¹⁸, de ser infinitamente sensibles a las condiciones iniciales. Al introducir cualquier perturbación en un punto de la superficie, aparecen allí nuevas ondas y estructuras de color.

Otra variante de este modelo, donde no hay reacciones sino solo difusión, sirve para explicar las celdas de Bénard (Figura 23) que se producen al calentar un líquido en una sartén. Visto lateralmente (Openlb, 2013), el flujo en cada celda es como un toroide descendiendo en el centro, enfriándose al llegar arriba y ascendiendo por los lados. Y visto desde arriba

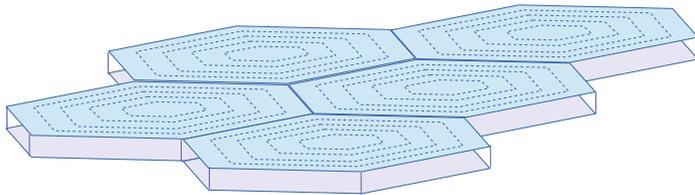
18 Recordemos que es imposible demostrar que haya caos en sistemas físicos, pues se requieren tres condiciones, dos de las cuales solo se pueden demostrar en sistemas formales, es decir, en modelos, no en sistemas físicos reales.

se forma un patrón de celdas más o menos hexagonales, que sugieren un posible origen matemático de los tejidos celulares (Guaicoloro, 2008).

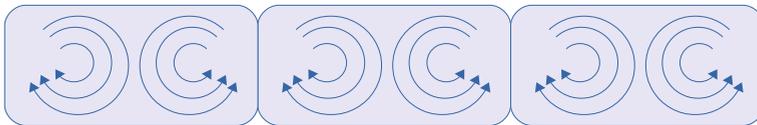


Figura 22. Reacción BZ. Fuente: CC BY 2.0, Michael Rogers (2010).

Disponible en: <https://goo.gl/gTSV2f>



(a)



(b)



(c)

Figura 23. Celdas de Bénard. (a) Patrón hexagonal; b) flujo del líquido, descendiendo en el centro de cada celda; c) fotos en una sartén con aceite y un medio de contraste

Se sabe que estas celdas existen en las altas temperaturas de la superficie del sol y, curiosamente, el mismo efecto (McKinnon *et ál.*, 2016) se acaba de observar incluso con las bajas temperaturas que hay en los mares de nitrógeno sólido de Plutón (Figura 24).

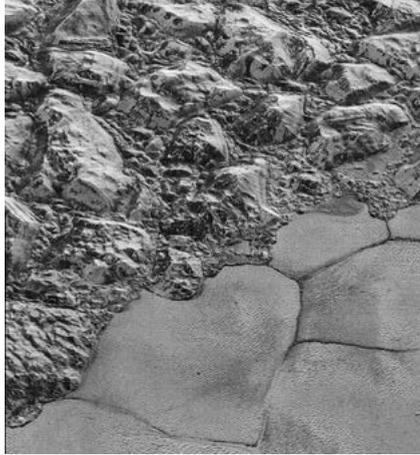


Figura 24. Mar de nitrógeno Sputnik Planum, en Plutón

Fuente: NASA/JHUAPL/SwRI

El modelo de Turing también explica la expansión de poblaciones de seres vivos, la propagación de enfermedades, etc., y se puede implementar en un espacio discreto 2D de tipo autómatas celulares. Como vemos, Turing fue otro de los grandes genios que no solo inventó los formalismos de la computación, sino los fundamentos de la vida artificial.

Otros ejemplos de autoorganización son las nubes (Figura 25) y ciertas formaciones rocosas (Figura 26).



Figura 25. Patrones en nubes, en la Chapada Diamantina, Brasil



Figura 26. Piedra de la Tortuga en Sete Cidades, Brasil.
Las grietas se formaron por contracción de la piedra al final del Devónico, lo que produjo unos patrones pentagonales y hexagonales

Las realimentaciones que autoorganizan un sistema suelen generar algún tipo de patrón, como en todas estas figuras.

Como explicábamos, se dice que ha emergido una nueva propiedad cuando de la interacción de muchos entes surge algo nuevo que antes no existía. En ese sentido la bandada es un fenómeno emergente que aparece cuando interactúan muchos pájaros. La emergencia es un fenómeno que se suele dar al superarse un umbral de número de entes que interactúan. Por debajo de un cierto número N no pasa nada, mientras que por encima aparece un nuevo fenómeno. Un ejemplo trivial puede ser un lobo solitario, que no caza nada. Pero una manada de lobos sí puede acorralar a su presa. Otro ejemplo en física nuclear: si juntas uranio, al principio emite calor. Si juntas más uranio, emitirá más calor. Pero si juntas mucho más uranio pasará algo nuevo: explotará. La masa crítica es 52 kg.

Estos “números mágicos” suponen una transición de fase al superar un umbral. Muchos de ellos requieren umbrales bastante elevados, del orden de 10^{10} :

- Aproximadamente 10^{10} átomos forman una hebra de ADN.
- Aproximadamente 10^{10} moléculas forman una célula.
- Aproximadamente 10^{10} células forman un órgano.
- Aproximadamente 10^{10} neuronas forman un cerebro. Por ejemplo, el comportamiento de una neurona aislada es bastante sencillo y existen modelos matemáticos para ello. Si conectamos unas pocas decenas de neuronas, el comportamiento del conjunto se puede deducir

fácilmente a partir del modelo matemático de la neurona aislada. Pero cuando se supera un umbral que debe de estar alrededor de 10^{10} neuronas interconectadas, aparecen comportamientos extremadamente complejos, como el pensamiento.

- Aproximadamente 10^{10} personas forman... ¿quién sabe? Se especula actualmente que la globalización, con todos los seres humanos conectados por Internet quizás esté generando un nuevo tipo de superorganismo. Algunos de los superorganismos más pequeños son fáciles de reconocer: son las ciudades. Todas tienen una estructura interna similar (centros de decisión, centros de compraventa de alimentos, redes de distribución de energía, agua y comunicaciones, redes de calles y lugares de esparcimiento). Es un patrón que se reproduce pues continuamente están naciendo nuevas ciudades. El patrón es el mismo, aun cuando las personas que constituyen cada ciudad sean distintas. Las ciudades primitivas crecían a lo loco, pero las modernas ya tienen algo de consciencia de sí mismas, pues poseen documentación como planos, historia, planeación futura, etc.
- Aproximadamente 10^{10} estrellas forman una galaxia.
- Aproximadamente hay 10^{10} galaxias en nuestro universo.

Este número que se repite tanto puede ser solo una casualidad, pues hay otros umbrales bastante menores en otros sistemas.

Los fenómenos emergentes comparten una serie de características:

- Ocurren cuando el número de objetos que interactúan supera un umbral, aun cuando no esté muy bien definido. El umbral a veces puede ser 3, como en el caso de la gravedad que con 2 cuerpos la teoría funciona perfectamente, mientras que con 3 o más cuerpos es un fenómeno complejo e impredecible. A veces pueden ser miles, como las hormigas incapaces de sobrevivir solas, pero que en un hormiguero se vuelven invencibles. A veces pueden ser miles de millones, como las neuronas que aisladas no hacen más que cálculos simples, pero que al superar el mencionado umbral producen inteligencia y consciencia.
- No hay un control centralizado. La hormiga reina es solo el aparato reproductor del hormiguero, pero ella no da órdenes. No hay un líder en la bandada de pájaros. Y dentro del cerebro humano no hay una superneurona que organice cómo deben trabajar las demás.
- Las interacciones entre objetos pueden ser muy sencillas, pero deben ser no lineales, para que el fenómeno emergente no se pueda predecir por simple superposición (suma) de muchas pequeñas interacciones.

- El fenómeno emergente nos pilla casi siempre por sorpresa. No es posible deducirlo de las propiedades de los objetos y sus interacciones. A veces incluso puede ser antiintuitivo, como cuando hay un banco muy sólido, pero aparece un rumor de que va a quebrar; ninguno de sus ahorradores desea que quiebre, pero cada uno de ellos, por temor, saca de allí su dinero, y entonces el banco efectivamente quiebra, contra todo pronóstico.
- El comportamiento que emerge no depende de ningún objeto en particular. Los objetos pueden tener una cierta variabilidad de características, no todos son iguales, pero cuando se agrupa una colección cualquiera de estos objetos, emerge el mismo fenómeno. Por ejemplo, en un río a veces se forman remolinos. El remolino es un patrón que permanece en el sitio, aunque las moléculas de agua que lo conforman entran y salen en un flujo continuo, es decir, son distintas en cada instante de tiempo. Esta idea está muy bien desarrollada en el libro *Sistemas Emergentes* de Steven Johnson¹⁹, donde también se explica que los barrios de ciertas ciudades se especializan en cierto tipo de comercio (Calle de Libreros, Rivera de Curtidores, etc.) y ese patrón permanece durante cientos de años a pesar de que la gente muere y nace entre medias. Con los hormigueros pasa algo similar, pues suelen durar alrededor de diez años, siendo que cada hormiga no suele superar los dos años de vida; y el hormiguero pasa por fases de juventud, donde es muy proclive a atacar a otros hormigueros, y madurez, donde prefiere mantenerse en paz alejado de problemas.
- El futuro de un sistema depende de su historia. La historia de un sistema complejo es irreplicable, por lo que poco se puede aprender de ella para predecir el futuro. Por ejemplo, la teoría de la evolución de Darwin explica muy bien cómo surgen las diferentes especies. Se puede hacer un análisis *a posteriori* de las mutaciones que dieron lugar a cambios en los seres vivos y se pueden trazar árboles filogenéticos. Pero, a pesar de conocer bastante bien el pasado, no podemos predecir que nuevos animales o plantas aparecerán en un futuro. Es más, si sembramos un planeta virgen similar al nuestro con las moléculas básicas de la vida, es imposible que se repita el mismo árbol filogenético de la Tierra. No hay garantías de que surjan dinosaurios y, si lo hacen, no hay garantías de que se extingan como ocurrió aquí para dar paso a las aves y permitir el amplio desarrollo de los mamíferos.

19 Que dice muy poéticamente que una ciudad, al igual que un remolino en el agua, es un patrón que se mantiene en el tiempo.

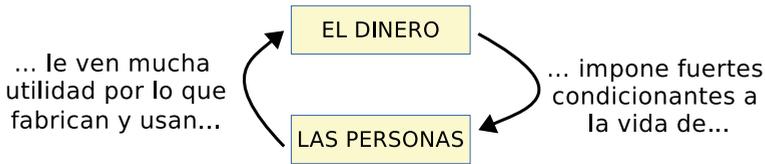


Figura 27. El dinero emerge de la interacción de las personas

Un ejemplo de emergencia más sofisticado es el dinero. Todos sabemos que el dinero es solo papel. Pero entonces, ¿por qué le damos tanta importancia? ¿Por qué no lo destruimos, en vez de ser sus esclavos? La respuesta es porque se trata de un sistema que se automantiene por realimentación positiva (Figura 27).

Podemos imaginar que el primer billete fabricado en el mundo nadie lo tomaría por otra cosa que un papel. Lo podríamos romper sin ninguna consecuencia. Pero conforme la gente lo va aceptando como algo valioso, se convierte en algo valioso. Es una realimentación positiva, porque cuantas más personas acepten el valor del dinero, todavía más personas se verán obligadas a hacerlo. A esto también se le llama una profecía autocumplida.

Una vez que el dinero es ampliamente aceptado por la mayoría de las personas, destruir mi dinero me pone en una situación de inferioridad frente a los demás. Por tanto, no lo voy a destruir y me toca imitar el comportamiento de la mayoría.

Además, aparecen instituciones creadas ex profeso, por y para el dinero, como los bancos y todo el sistema financiero. En este caso, la emergencia ya no solo es una cuestión subjetiva de reinterpretación del mundo, puesto que estos nuevos objetos²⁰ solo tienen sentido dentro del mundo que está naciendo.

Algo similar ocurre con la visión *newtoniana* de la gravedad, donde cada masa interactúa directamente con cada otra, comparada con la visión relativista de la gravedad, donde el espacio juega un nuevo papel como intermediario: las masas modifican la curvatura del espacio, y este modifica las trayectorias de las masas. Por cierto que esta última versión de la gravedad podría ser sustituida por otra de la Física Digital²¹, propuesta por Verlinde, que explica la gravedad como un fenómeno emergente a partir de la entropía, y que eliminaría de nuevo ese objeto intermedio (el espacio). Hay un blog de ciencia en español, Naukas, donde están los enlaces a los trabajos más importantes en este sentido (Villatoro, 2010).

20 Aunque siempre podemos pensar que los nuevos objetos se apoyan mutuamente en mostrar su realidad, de manera similar a como las palabras de un diccionario se definen mutuamente unas a otras.

21 *Digital Physics*.

Otro ejemplo similar, en Física Digital: según la teoría de cuerdas, el electrón dentro del átomo de hidrógeno es una vibración que interfiere constructivamente consigo misma. Las demás vibraciones se autoanulan porque son equivalentes a ondas que interfieren destructivamente. El electrón genera la única onda posible en una cierta órbita, y esa onda genera el electrón. Es una emergencia a partir de una realimentación positiva.

Otro ejemplo de emergencia ocurre cuando tu computador personal se bloquea. En el sistema operativo no existe una instrucción que diga:

`tiempo_que_falta_para_el_siguiete_bloqueo = 3204; # segundos`

El bloqueo ocurre por la interacción de programas con el sistema operativo, cada uno con distintos objetivos, que requieren más o menos recursos, memoria, o intercambio con la memoria virtual. No es fácil de prever cuándo ocurrirá o las razones exactas para ello, pero si se puede averiguar experimentalmente dónde está el umbral.

Prácticamente todo lo que existe es un fenómeno emergente de una u otra forma. Desde los átomos (que emergen a partir de hadrones y leptones) hasta los derechos humanos y sindicales (que solo aparecen cuando hay un nivel de riqueza y confort suficientes para permitir reflexionar y llegar a acuerdos, más allá de la mera supervivencia diaria). Cuando estás en un nivel previo a los átomos (la sopa de quarks y gluones que se supone se produjo después del Big Bang) es imposible predecir que ellos van a aparecer. Y los derechos humanos podrían desaparecer si el nivel anterior que los sustenta se vuelve demasiado frágil.

En el mencionado libro de Steven Johnson se analizan muchos más ejemplos, incluyendo la interacción de personas que conforman barrios y ciudades, que están vivos en cierto modo; la interacción de hormigas que producen un superorganismo que llamamos hormiguero; la interacción de neuronas que producen un cerebro con pensamientos, inteligencia y consciencia; y la interacción entre una gran cantidad de humanos a través de Internet que, según Johnson no produce nada porque en la época en que escribió el libro los enlaces web eran unidireccionales lo cual impide crear bucles de realimentación. Eso ya ha cambiado y hay realimentaciones por todos los lados. Por ejemplo, las páginas web más populares aparecen mejor *rankeadas* en los buscadores por lo que más gente todavía llegará a ellas. Los sistemas de calificación *peer-to-peer* en los blogs permiten ofrecer nuestra opinión en distintos tópicos y evaluar y ser evaluados por otros. En las redes sociales las personas con más contactos están más cerca de muchas otras personas con las que también pueden establecer contacto. Eso nos permite intuir que las relaciones entre tantas personas mediadas por Internet pueden dar lugar a la emergencia de un nuevo superorganismo,

o quizás varios, y seguramente ello ya ha ocurrido, pero ¿puede una célula percibir que existe un cuerpo al que pertenece?



Figura 28. Moho del fango en una finca cafetera del Quindío, Colombia

Por cierto que el libro de Johnson comienza analizando un organismo muy peculiar, el moho del fango²², que suele vivir como células aisladas la mayoría del tiempo, pero que cuando el entorno es adverso se agrupa y forma un cuerpo. Emerge un organismo pluricelular (Figura 28).

Este organismo se ha puesto muy de moda pues es capaz de resolver problemas de búsqueda de rutas óptimas, que sabemos que son de complejidad NP (Hanson, 2017). Lo hace sin tener un control central, usando reacciones químicas como sistema de comunicación entre sus células. El moho del fango logra encontrar la salida a laberintos porque consiste en muchísimas células que se comunican entre sí por medios químicos. Cada célula no tiene casi inteligencia, sino que esta emerge del conjunto.

Hay otra forma de salir de un laberinto: inundándolo de agua. Me subo a un flotador y dejo que la corriente de agua me arrastre hacia la salida, que es por donde el agua escapará. Si hay un camino largo pero cerrado al final, es decir, una especie de óptimo local, el agua no fluirá hacia allá de manera sostenida. O sea que el agua me sirve para solucionar un problema a pesar de que no puedo asignar ninguna inteligencia a sus moléculas. Podemos decir que la inteligencia es un fenómeno colectivo. Analizándolo un poco más vemos que la interacción entre moléculas de agua es por choques junto con la gravedad que guía todo el proceso. Usamos el mundo físico como

22 *Slime mold.*

computador, de la misma manera que veremos en el siguiente capítulo que lo hacía Gaudí, para calcular la estructura de su catedral. O como estamos usando los nuevos computadores cuánticos.

Otro caso de inteligencia colectiva lo viví como director de programa. Resulta que los reglamentos de la universidad contienen inconsistencias y vacíos. Y dado que hay muchos estudiantes, terminan encontrándolos, para sorpresa de los altos mandos. Es similar a pensar en un barco que tiene un pequeño agujero. Me explican que el agujero está en un sitio muy complicado, en una zona de difícil acceso y que prácticamente no se ve, por lo que es muy improbable que alguna molécula de agua lo encuentre. Disiento. Y el barco, obviamente, se irá a pique. Cuando hay muchos estudiantes pasa lo mismo.

Hay infinidad de casos donde surge algo imprevisto de la interacción entre individuos. Vamos a ver ahora el modelo de segregación de Schelling, que muestra emergencia, en este caso de algo tan malo como la discriminación social, a pesar de que todos los individuos que interactúan son bastante amigables. En Hart (2017) se puede jugar con una versión interactiva de este modelo, que consiste en una cuadrícula. En cada casilla puede haber un individuo o puede estar vacía. Hay dos tipos de individuo (por ejemplo, de dos estratos sociales distintos, razas o equipos de fútbol) y todos ellos tienen una preferencia de vecindad. Para entendernos, si la preferencia es 20% eso significa que cada individuo mira a sus vecinos inmediatos y es feliz si al menos el 20% son del mismo tipo que él. Los agentes que no son felices se cambian a alguna de las casillas vacías. El proceso se repite hasta que todos los individuos sean felices.

Este porcentaje de preferencia de vecindad es común a toda la población y, para aclarar las ideas, mencionemos que una preferencia de vecindad del 50% significa que todos los individuos son felices si hay una mezcla homogénea de tipos de individuo, es decir, si la mitad de mis vecinos son como yo y la otra mitad son distintos a mí. Es decir, ese es el ideal de una sociedad sin segregación. Sin embargo, por paradójico que pueda parecer, la dinámica del modelo amplifica la segregación. Incluso con porcentajes de preferencia de vecindad tan bajos como el 30% (es decir, soy tan abierto que admito a un 70% de vecinos distintos a mí) y con una población colocada al azar en la cuadrícula, después de iterar aparecen grandes zonas de un tipo de individuos, aisladas de otras grandes zonas del otro tipo, alcanzándose valores de segregación totales del 56%. Y si partimos de un 50% de segregación en el individuo, se alcanza hasta un 80% en el conjunto de la sociedad. El fenómeno es parecido al de la langosta que hablábamos antes, pues cuando un vecino cambia de sitio conduce a que otros vecinos también cambien.

Para ilustrar lo que es una transición de fase se recurre muchas veces a los relojes de manillas que actualmente, giran en sentido dextrógiro (hacia la derecha cuando pasan por arriba). En los primeros modelos de los siglos X al XIII existían las dos versiones, dextrógiros y levógiros. Sus precursores, los relojes de sol, también existían en las dos versiones que cambiaban de sentido según fueran de pared (verticales, dextrógiros) o de mesa (horizontales, levógiros). En el hemisferio sur de la Tierra se comportaban exactamente al revés. Ambas versiones eran igual de buenas y exactas. ¿Por qué finalmente triunfaron los dextrógiros? La respuesta es decepcionantemente simple: por azar. Una vez que aparece una pequeña desviación en la cantidad de relojes de uno u otro tipo, se autoamplifica como un fenómeno de realimentación positiva: cuanta más gente aprendía a leer un reloj dextrógiro enseñaban a otra gente a hacerlo así. Cuantos más relojeros aprendían a fabricar y reparar relojes de un tipo, más relojes surgirían de ese tipo, lo cual produce una transición exponencial. Algunas de las piezas, como los resortes y los trinquetes, no son intercambiables entre los dos tipos de relojes sin hacer grandes cambios al resto del mecanismo, de modo que una vez que aparece una tendencia, es más económico unirse a ella. Los físicos lo llaman ruptura de simetría y los cibernéticos, bifurcaciones. Las condiciones iniciales permiten varias posibilidades, pero una vez que por azar se ha decidido avanzar hacia una, es muy costoso dar reversa para intentar llegar a la otra.

Aunque la emergencia siempre nos sorprende, se pueden identificar en abstracto algunos mecanismos que la producen y que, en el fondo, son el mismo:

- La emergencia se produce porque aparece un bucle de realimentación positiva que lleva al sistema a un comportamiento nuevo. Después de un tiempo, algo se agota y un bucle de realimentación negativa toma el control, estabilizando al sistema. Ha aparecido algo nuevo y se ha estabilizado. (Por cierto, si no aparece el bucle negativo, el sistema morirá).
- La emergencia se produce como consecuencia de procesos acumulativos no lineales (caóticos). En sistemas lineales se da que $f(a+b) = f(a) + f(b)$. En sistemas no lineales, aparecen términos nuevos. Las ecuaciones suelen ser no resolubles o no computables en tiempo finito, con lo que el resultado final siempre es una sorpresa.

Cada vez que emerge un nuevo nivel ontológico, lo hace porque una realimentación negativa lo estabiliza. Desde allí puede emerger otro nivel, y desde allí otro, de manera indefinida, hasta que los recursos se agoten. Pero solo uno de ellos tiene libertad (en ese nivel es que ocurre “la vida”) y es

habitualmente el último (aunque no siempre²³) pues no se puede construir algo si la estructura de la que se parte es imposible de predecir. En el capítulo “Libertad” hablaremos un poco más de ello.

Si el nivel actual es determinista, se pueden usar las “leyes” que ofrece para construir otro nivel por encima. Incluso si es aleatorio donde se pueden sacar promedios, también se puede generar otro nivel de complejidad superior. Pero si es aleatorio “salvaje”, no hay ninguna posibilidad de predecirlo y, por tanto, no puede emerger nada estable basado en ese nivel.

Actualmente somos muchos seres humanos y tenemos muchas conexiones gracias a Internet, por lo que se especula²⁴ que ya debe estar emergiendo el siguiente nivel, al que podemos llamar algo así como *Gaia Inteligente*. Y para que surja como entidad independiente, necesita que los niveles inferiores (nosotros, los humanos) funcionen de forma predecible, o sea, sin libertad. Por ello se puede especular que la consciencia de los humanos va ser un fenómeno transitorio ya que es una realimentación que nos da libertad (miro lo que hago y tomo decisiones para hacer cambios en comportamientos futuros, como veremos en detalle en el capítulo sobre la consciencia). Un sistema sin realimentaciones es muy predecible, y con realimentaciones deja de serlo, de manera que nuestra libertad y consciencia humanas son un obstáculo para que se desarrolle *Gaia Inteligente*.

GENERADOR DE COMPLEJIDAD EVOLUTIVO

Habitualmente los algoritmos evolutivos (como los presentados en el libro anterior) se han usado para resolver problemas de ingeniería y de diseño, utilizando la aptitud para guiar la evolución hacia una solución con ciertas características deseadas. Pero también se pueden usar para generar complejidad, como veremos a lo largo de este capítulo.

Una sorprendente limitación de los algoritmos evolutivos tradicionales es que como la función de aptitud es fija y se pone de antemano para guiar a una solución, la emergencia (de soluciones compatibles con lo que buscamos) ocurre solo una vez. Por ejemplo, si queremos diseñar una silla resistente, económica y ergonómica entonces partimos de una población de

23 Por ejemplo, Dennett (1999) explica que en un autómata celular se puede implementar un algoritmo distribuido que juegue al ajedrez. A alto nivel, si le ofrezco mi reina a cambio de un peón, sé que el algoritmo se comerá la reina. Mientras que a bajo nivel el autómata celular es muy difícil de predecir a largo plazo. Lo habitual es al revés (el alto nivel tiene libertad, y el bajo nivel no), pero lo general es que solo uno de los niveles tenga libertad.

24 No solo en ciencia ficción como la *Skynet* de *Terminator* y Clarke (1979).

cromosomas generados al azar que codifican posibles sillas, con materiales, número y disposición de patas, y forma del respaldo y asiento. Las primeras sillas son un desastre debido a que son aleatorias (por ejemplo, tendrán 7 patas, algunas hacia arriba o con el asiento en vertical). Esas sillas tendrán muy baja funcionalidad y ergonomía, de modo que su aptitud será pequeña y tendrán pocos hijos. Con el paso del tiempo, las sillas que por mutación o cruce descubran formas más afortunadas, tendrán más hijos. Y al final podremos decir que emergió un nuevo objeto, la silla, con las propiedades que buscamos. Quizás sean sillas vulgares de 4 patas, asiento y respaldo o quizás la evolución encuentre otras formas más cómodas y funcionales. Pero en cualquier caso emerge un objeto útil. Desgraciadamente, si dejamos la evolución en marcha más tiempo no va a emerger nada más. No nos va a sorprender con un sillón, un sofá o un sofá-cama, porque en la especificación de la aptitud deseable no hemos puesto que sea mullido (que nos llevaría al sillón), que pueda ser ocupado por más de una persona (que nos llevaría al sofá) o que pudiera tener otras funciones extra (que nos llevaría al sofá-cama). Los requerimientos que no estén explícitos en la función de aptitud nunca se van a tener en cuenta. Y si los ponemos todos de una vez obtendremos de nuevo un único nivel de emergencia (el sofá-cama) sin pasar por niveles de emergencia intermedios ni soñar con que pueda producir niveles de emergencia en un futuro (como un diván, un colchón de agua, una silla para automóvil, etc.).

La emergencia solo ocurre una vez. Pero eso no es lo que vemos en los sistemas biológicos. ¿Qué es lo que falla entonces en nuestros algoritmos evolutivos? Se han hecho dos propuestas para superar esta limitación: la coevolución (Goldberg, 1989; Mitchell, 1999) y la supresión de la aptitud (EPlex Group, 2017).

Tenemos explicada la coevolución en el libro anterior. Recordemos que la idea principal es aumentar poco a poco la complejidad del problema, para que la solución buscada también lo haga. En el último ejemplo primero pediríamos (calculando adecuadamente la aptitud) que la silla sea funcional (que no se rompa, que no se caiga) y cuando la población de sillas alcance este requerimiento entonces pediríamos que también sea ergonómica (adaptada a las medidas de una persona), y luego que también sea cómoda (empleando materiales mullidos). Esto es laborioso, requiere mucha intervención del programador humano y, por tanto, es poco general, poco práctico.

La otra posibilidad, suprimir la aptitud, consiste en dejar grandes periodos de tiempo donde la evolución solo genere variedad, sin ninguna presión selectiva orientada a cumplir requerimientos. A lo sumo se puede considerar una aptitud que reproduzca más a los cromosomas más raros, más

diversos, más distintos al promedio. Una vez que se tiene una población muy variada se continúa la evolución con la aptitud tradicional, orientada a satisfacer requerimientos.

Y hay una tercera propuesta que es mixta entre las dos anteriores: coevolución sin una finalidad explícita. Por ejemplo, podemos modelar predadores y presas en un espacio plano sobre el que se pueden mover. Si un predador y una presa se encuentran en las mismas coordenadas espaciales, el primero devora al segundo. Y si un predador permanece mucho tiempo sin comer, entonces muere. Por ello, aunque no esté explícitamente codificado en el programa, el objetivo de los predadores será acercarse a las presas para comérselas, y el objetivo de las presas será huir de los depredadores. Esto dará lugar a una serie de emergencia de estrategias de los predadores y de las presas, para cumplir con sus objetivos implícitos.

Sin embargo, todos estos métodos tienen una gran limitación que quiero ilustrar con los dos ejemplos usados: una sala de teatro es, básicamente, un conjunto de sillas dispuestas de cierta forma; esto jamás emergerá de un algoritmo evolutivo tradicional, ni con las variantes que acabamos de explicar, debido a que la estructura del cromosoma es fija y no permite ir más allá de patas, asiento y respaldo. Y respecto a los predadores y presas, a la presa jamás se le ocurrirá generar un par de alas y huir volando porque en su cromosoma no están codificadas; podrá correr en zig-zag, alargar el tamaño de las patas y cosas similares siempre y cuando tenga genes que lo codifiquen. Pero las alas son imposibles porque no existen esos genes inicialmente.

Queda así identificado el problema: los cromosomas son diseñados por el programador como una secuencia fija de genes que codifican cosas relevantes a la solución buscada. Pero son fijos. Si al programador se le olvida algo, no hay nada que hacer. El algoritmo evolutivo no puede inventar un gen para las alas de la presa. La naturaleza es dinámica mientras que los cromosomas de nuestros algoritmos son estáticos.

Hay varias formas de solucionar este problema, pero una muy interesante y general está descrita en el libro *Ocho hitos de la evolución* de John Maynard Smith y Eörs Szathmáry. No me entiendan mal, pues se trata de un libro sobre biología, no sobre programación. Pero lo que Maynard y Szathmáry describen en cada uno de los hitos es exactamente el mismo algoritmo, si uno sabe interpretarlo adecuadamente. Veamos cómo es.

Su idea es que la evolución consiste principalmente en un proceso competitivo tal y como explica otro neodarwinista como Dawkins en *El gen egoísta*. Esto casa perfectamente con el concepto de optimización que empleamos en ingeniería: las garras del predador se van haciendo más afiladas y las patas de la presa más rápidas conforme pasa el tiempo, porque con ello

mejoran sus capacidades de supervivencia, con lo que tendrán más hijos con cualidades similares respectivamente, lo cual lleva a un proceso de mejora paulatino.

Sin embargo, Maynard y Szathmáry identifican también ocho saltos discontinuos en ese proceso de mejora continua (que, por cierto, se corresponden a la idea de equilibrio puntuado de Niles Eldredge y Stephen Jay Gould, y que había generado tantas polémicas entre biólogos gradualistas y saltacionistas, ambos evolutivos). Estos saltos son:

- De moléculas replicantes aisladas, a encerradas dentro de membranas.
- De replicadores independientes, a asociados en un mismo cromosoma.
- Del mismo *ARN* actuando como gen y encima, a la separación de su trabajo en *ADN* y proteína.
- De la célula procariota, a la célula eucariota.
- De la reproducción asexual (clonado de una única célula), a la reproducción sexual (fusión de dos células).
- De los protistas (individuos unicelulares), a los animales, plantas y hongos (individuos pluricelulares).
- De los individuos solitarios, a las sociedades de individuos.
- De las sociedades de primates, a las sociedades de humanos por medio de la aparición del lenguaje.

El libro ya tiene 20 años y hay experimentos nuevos que corroboran y afinan con más precisión todo ello. Pero casi todas estas transiciones tienen un algoritmo común: se lograron por la cooperación de los individuos del nivel anterior. Los casos triviales de entender son el paso de seres unicelulares a multicelulares o el paso de individuos aislados a individuos que viven en sociedad. Pero también las células eucariotas se formaron por asociación de bacterias (y prueba de ello es que estas se convirtieron en estructuras internas de aquellas, como los cloroplastos, núcleo, membrana, mitocondrias etc. y muchas de ellas todavía conservan vestigios de sus propios genes). Lynn Margulis (1998) lo explica muy bien en su libro *¿Qué es el sexo?*

El algoritmo que descubrieron Maynard y Szathmáry alterna momentos de competencia, donde se optimizan las funcionalidades de los individuos, con momentos de cooperación, donde los individuos se asocian para formar algo más complejo. Es gracias a la cooperación que emergen nuevos niveles de individuos formados por asociación de los del nivel anterior. El proceso es iterativo.

El algoritmo que propongo para modelar este proceso consiste en los siguientes cinco pasos:

1. Autocopiado. Un individuo logra la capacidad de sacar copias de sí mismo.
2. Competencia. Gracias al autocopiado, aparece una familia de individuos muy parecidos pero no idénticos, pues siempre habrá pequeños errores en las copias que hacen que la población evolucione. Como usan los mismos recursos, habrá un proceso de competencia que produce cada vez mejores funcionalidades.
3. Agrupación. Por accidente aparece algún tipo de membrana o pegamento o ligazón que los agrupa en un “superorganismo”. Puede ser que uno quede dentro de otro, que un fenómeno físico los aisle, etc.
4. Cooperación. La nueva unidad solo se reproduce y vive con ventajas frente a otros si todos sus individuos internos colaboran.
5. Especialización. Inicialmente todos los individuos son iguales dentro del superorganismo. Pero con el paso del tiempo, las mutaciones (errores de copia) hace que se diferencien. Es ventajoso que se especialicen en hacer cada uno una cosa diferente, pero siguen colaborando. De hecho, ya no pueden dejar de colaborar, pues dependen físicamente unos de otros.

Para poder iterar, regresando al paso número 1, se requiere que el superorganismo retenga su capacidad de autocopiado, quizás porque sus individuos se siguen replicando y luego vuelven a asociarse, o quizás porque descubre una nueva forma de replicarse como un todo.

Los algoritmos evolutivos clásicos implementan únicamente los pasos 1 y 2 que constituyen lo que se conoce como evolución gradualista. El paso 3 de agrupación es el que fabrica individuos con nuevas estructuras genéticas. El paso 4 de cooperación podría no darse, pues es un fenómeno difícil (y seguramente por eso, a lo largo de los 4.100 millones de años de vida en la Tierra, solo se ha dado esas ocho veces). En el capítulo “Vida” estudiaremos la cooperación más detenidamente. Pero si se da la cooperación, el paso 5 de especialización es inevitable, pues las mutaciones internas producen individuos ligeramente distintos, y la evolución a nivel superior presionará para que cada uno haga lo que sabe hacer lo mejor posible. Es decir, se especializará. Y una vez que ello ocurra, no tiene vuelta atrás, porque se producen dependencias mutuas. Por dar un ejemplo con humanos viviendo en sociedad: nada impide que un individuo fabrique sus zapatos, sus muebles, coseche su comida y levante su casa, como Robinson Crusoe. Pero ello es costosísimo en tiempo, dinero y calidad. Es mucho más eficiente que el zapatero fabrique zapatos, el carpintero los muebles, el panadero pan, el arquitecto levante las casas, y todos intercambien el resultado de su trabajo

con los demás. Es costoso que cada persona tenga todas las herramientas del carpintero, del panadero y del arquitecto. Es costoso intelectualmente que sepa usarlas bien. Y es dudoso que tenga tiempo para hacer tantas cosas.

Además, como decíamos, una vez que surge la especialización no suele tener reversa porque cada ente es incapaz de realizar todos los procesos que se requiere para mantenerse vivo, ya que delega ciertas funciones en otros entes. Este es uno de los trinquetes²⁵ de la evolución²⁶. Por cierto que el otro trinquete de la evolución es la propia complejidad que genera: una mutación aislada es aproximadamente igual de probable que ocurra en un sentido o en otro, creando o destruyendo cierta funcionalidad. Pero cuando ocurre una mutación afortunada suele abrirse un nuevo espacio de diseño que favorece muy rápidamente una cascada de mutaciones adicionales. Deshacer tantos cambios y tan interdependientes es muy improbable y costoso. Por poner una metáfora, si decides cambiar de domicilio ello favorece también que compres nuevos muebles adaptados a las dimensiones de tu nueva casa, hacer nuevos amigos en el barrio, modifiques tus rutinas de compras y de paseo, etc., con lo que se complica muchísimo deshacer la acción inicial (en el sentido de que la evolución no elige rutas costosas).

Por último, que el organismo retenga la capacidad de autocopiado no es obvio, aunque existen varias formas de lograrlo, que se analizarán en el capítulo “Vida”.

Este algoritmo tiene la ventaja de que su futuro está abierto, muy al contrario de los algoritmos evolutivos tradicionales. Pueden emerger sucesivos niveles de complejidad cada vez que se repiten las fases de 1 a 5. Y la complejidad que genera no es predecible, principalmente porque en la fase de agrupamiento hay muchas formas de realizarlo. Es un proceso combinatorio. Por ejemplo, no es necesario que todos los entes que se agrupen sean homólogos, ni siquiera del mismo nivel: un retrovirus puede “agruparse” con un mamífero; una persona puede prestar servicios a una empresa sin pertenecer a ella; las células eucariotas probablemente se formaron como simbioses entre una célula metanogénica y una eubacteria, que finalmente desapareció en las arqueobacterias, o se transformó en hidrogenasa o en mitocondria en las eucariotas; todo el reino *Chromista*, donde cada individuo es una simbiosis de varios otros, cada uno con su propio ADN; la nanomia, que es un animal de mar formado como colonia a partir de varios animales distintos.

25 Mecanismo de engranajes que permite girar o transmitir fuerza en una dirección pero no en otra, por ejemplo, la transmisión de los pedales a la rueda trasera de la bicicleta. En inglés “*ratchet*”.

26 Por eso es que el Brexit es tan complejo, porque ya hay muchas dependencias mutuas difíciles de cortar.

RESUMEN

En este capítulo proponemos una forma de ver el mundo, a través de su complejidad. Conforme aumenta la complejidad aparecen fenómenos (inteligencia, consciencia, libertad e incluso realidad) que no son de todo o nada, sino que hay una cierta gradualidad en su expresión. A partir de ahora, en sucesivos capítulos, nos vamos a centrar en estudiar la libertad, la capacidad de cómputo universal, la autoorganización, la capacidad de sacar copias, la inteligencia y la consciencia.

Hemos visto algunas formas de medir la complejidad de un objeto, para lo cual hay que entresacar lo estructural y rechazar lo que dependa del azar. Todas las medidas de complejidad arrojan únicamente cotas superiores, pues siempre cabe dentro de lo posible comprimir más la información estructural. Con ello hemos visto la definición de bit, y que los físicos comienzan a considerar el mundo formado no solo por materia y energía, sino también de información.

Hemos hablado de las formas como el mundo genera complejidad a partir de muy poco. Los siguientes mecanismos generan complejidad a partir de la nada, o más bien, de la casi-nada:

- La referencia (que los objetos se puedan diferenciar) y la autorreferencia (un sistema que pueda decir cosas sobre sí mismo) aumenta muchísimo la complejidad porque se trata de una realimentación positiva, como se puede observar en el teorema de Gödel, los sistemas autorreferentes de Douglas Hofstadter y los fractales autosemejantes.
- La combinatoria en la agrupación de objetos. Una vez que los objetos agrupados interactúan entre sí, es trivial que aparezcan realimentaciones. Con la teoría de juegos se pueden estudiar las formas de interrelación.
- La realimentación (son las matemáticas más difíciles de las ingenierías). Suelen producir: caos (los sistemas de ecuaciones no lineales que producen caos no suelen tener solución analítica); fractales (curvas autosimilares); y leyes de potencias (distribución de probabilidades donde los eventos mayores ocurren pocas veces y los eventos menores ocurren muchas veces).
- La evolución para lo cual se requiere que los objetos puedan sacar copias de sí mismos. Y de los objetos que evolucionan puede emerger la cooperación.

Luego hemos hablado que la complejidad puede aumentar gradualmente (como cualquier proceso de optimización por búsqueda local), pero también

puede dar saltos (típicamente por procesos de realimentación positiva que generan exponenciales) e incluso puede disminuir (y hemos explicado cuando puede ocurrir esto).

Al final, propongo un algoritmo general para crear complejidad.

Quiero también dar una advertencia: no sabemos cómo funcionan los sistemas complejos. Es difícil predecir y controlar su comportamiento, así que debemos ser cautos. No hay nada tan peligroso como un burro con iniciativa. Por ejemplo, para aumentar el control sobre un sistema es habitual (pero equivocado) aumentar las restricciones (normas, leyes...). Esto suele ser un error ya que las restricciones interactúan entre sí de forma combinatoria, con lo cual aumenta la complejidad e impredecibilidad del sistema. Además, al aumentar el número de restricciones (normas, leyes) por encima de un cierto nivel, es casi seguro que se producirán contradicciones entre ellas. Un individuo hábil puede lograr cualquier cosa de un sistema así. En ese caso observaremos la emergencia de algo nuevo e inesperado. No olvidar que los sistemas complejos son realmente complejos.

Para saber más

En los últimos años se ha escrito mucho y de mucha calidad sobre complejidad. Recomiendo leer los siguientes libros, la mayoría de los cuales son muy amenos.

- Steven Johnson (2003). *Sistemas emergentes*. México: Fondo de Cultura Económica.

Es un libro espectacular. Su tesis principal es que una hormiga en un hormiguero hace lo mismo que una neurona en un cerebro y que una persona en una ciudad. Si lo lees, cambiará tu mente. En el último capítulo explora si Internet es un sistema suficientemente complejo para poder decir que tiene vida propia y este es el único aspecto donde se queda un poco corto.

- James E. Lovelock (2000). *Las edades de Gaia*. Barcelona: Tusquets Editores.

El libro es fascinante. Cuenta cómo es posible detectar si hay vida en un planeta examinando los gases de su atmósfera. Hay muchos datos biológicos interesantísimos y poco conocidos, sobre procesos de autorregulación del planeta Tierra. Entre ellos, está el *software* DaisyWorld como modelo para explicar cómo se ha mantenido constante la temperatura de nuestro planeta. Explica la verdadera razón de que el agua de los planetas desaparezca con el tiempo (y no, no es porque te dejes la llave abierta, pues eso solo genera un problema económico, pero no ecológico en primera instancia). Cuenta

que encontraron un reactor nuclear natural (biológico) y la historia de que el oxígeno es un veneno peligroso. Muchos problemas actuales dejarían de serlo si se mantiene la población humana por debajo de 500 millones.

- John Maynard Smith y Eörs Szathmáry (2001). *Ocho hitos en la evolución. Del origen de la vida a la aparición del lenguaje*. Barcelona: Tusquets Editores.

La evolución procede lentamente, pero a veces da saltos que producen cambios drásticos (nuevas estructuras con mayor complejidad). Los autores muestran cuáles han sido esos cambios en biología. Es un libro muy sugerente para quien pretende abordar el problema de la emergencia de la complejidad desde el punto de vista computacional. Requiere unas bases mínimas de biología.

- William Irwin Thompson, Gregory Bateson, Francisco Varela, Humberto Maturana, James Lovelock, Lynn Margulis, Henry Atlan, John Todd y Hazel Henderson (2001). *Gaia: uma teoria do conhecimento*. São Paulo: Editora Gaia Ltda.

Un libro que reúne a los más sobresalientes investigadores en el tema de los sistemas complejos. Analiza muchos sistemas ecológicos usando la teoría de la complejidad. Analiza también problemas de política y de economía (principalmente la globalización), como sistemas complejos. Es un libro extraordinario pues su primera edición es de 1987 y allí se predicen cosas que vemos hoy día cumplidas. Está en portugués, pero hay versiones en inglés y en español.

- José Navarro Cid (2001). *Las organizaciones como sistemas abiertos alejados del equilibrio*. [Tesis Doctoral]. Barcelona: Facultad de Psicología de la Universidad de Barcelona.

Una aplicación de la complejidad y la emergencia a problemas de organizaciones humanas. La primera parte contiene una muy clara y valiosa introducción a los conceptos de teoría general de sistemas, teoría de la complejidad, realimentaciones positivas y negativas, retardos, emergencia, complejidad, caos y fractales. Cita los trabajos de Stuart Kauffman, Mandelbrot, Prigogine, Maturana y Varela.

- M. Mitchell Waldrop (1992). *Complexity*. New York: Simon & Schuster.

Da una descripción histórica de lo que fueron los primeros desarrollos científicos modernos en el área de la complejidad. El libro ha quedado anticuado por cuanto el área se ha desarrollado mucho, pero tiene interés por-

que plantea las preguntas básicas, y cuenta las historias de las personas que hubo detrás. Es parecido al libro de Lewin.

- Roger Lewin (1992). *Complexity: Life at the edge of chaos*. New York: MacMillan Publishing Company.

También da una descripción histórica de lo que fueron los primeros desarrollos científicos modernos en el área de la complejidad. El libro ha quedado anticuado por cuanto el área se ha desarrollado mucho, pero tiene interés porque plantea las preguntas básicas, y cuenta las historias de las personas que hubo detrás. Es muy parecido al libro de Waldrop, y para mí es un misterio por qué se escribieron a la vez dos libros tan similares.

- Murray Gell-Mann (1996). *El Quark y el Jaguar*. Barcelona: Tusquets Editores.

Gell-Mann, premio nobel de física, fue uno de los pioneros en abordar el problema de la complejidad, esto es: ¿cómo ocurre que la reunión de muchas cosas simples (como los quarks) pueden dar lugar a algo increíblemente complejo (como un jaguar)? Fue el primer director del Instituto de Santa Fe, uno de los primeros centros de investigación en el tema de la complejidad.

- Stuart A. Kauffman (2000). *Investigations*. New York: Oxford University Press.

El libro es altamente especulativo, pero muy sugestivo. Se requiere una cuarta ley de la termodinámica para explicar la vida, y el autor trata de aproximar cómo podría ser esa ley, intentando explicar la creatividad del universo, la emergencia de nuevos entes y propiedades. Muestra una perspectiva donde todas las leyes del universo, y el mismo espacio-tiempo, son resultado de procesos evolutivos (Smolin). Habla de los teoremas de *No-Free-Lunch* y del *halting problem* aplicado a la química (no tiene nada de raro, pues la química también son algoritmos). Habla de leyes de potencias (fractales), criticalidad autoorganizada, coevolución, astrobiología, *SETI* y una nueva área: la teoría general de la biología. Comenta la famosa cita a Huxley: “la cosa más extraña de la teoría de la evolución es que todo el mundo cree que la entiende”. Filósofa sobre la ciencia, citando a Poncaire, Wittgenstein, Pierce, Dennett, Hume, Pareto, Gödel, Feynmann, Smolin, Rawls (con su justicia evolutiva). Presenta el *software ALCHEMY* de Walter Fontana donde pueden observarse 3 niveles de emergencia. Utiliza una idea suya, las redes booleanas estocásticas, para simular genes. Comenta el trabajo de Bak sobre criticalidad autoorganizada en montones de arena. Muestra que el problema *KSat* (que es NP) tiene una transición de fase. Da una idea brillan-

te de cómo resolver el *job shop scheduling* manteniendo el problema en estado de percolación (en el filo del caos), que usamos en EVALAB en el 2009, en el trabajo de grado de la estudiante María Alejandra Céspedes. Muestra, al igual que tantos otros, que la economía es un proceso evolutivo. Habla de la teoría de los *spin networks* de Roger Penrose, para mostrar cómo puede crearse el espacio-tiempo. Algunas páginas (pocas) deben de saltarlas los estudiantes de informática, pues están escritas en un lenguaje para químicos, impenetrable. De la página 254 en adelante no merece la pena ser leído por un estudiante de informática puesto que explica de manera muy superficial y atropellada la teoría de cuerdas y de *M-branas*. Hay cierta intersección entre los todos los libros de Kauffman, pero merecen la pena ser leídos.

- Stuart A. Kauffman (1993). *The origins of order. Self-organization selection in evolution*. New York: Oxford University Press.

La evolución no puede explicar satisfactoriamente cómo apareció la primera molécula autorreplicante. El autor nos muestra que el proceso es matemático (lo llama autoorganización en el filo del caos) y ocurre inevitablemente. Habla de los paisajes adaptativos rugosos, que se encuentran cuando empleamos algoritmos evolutivos. Muestra la similitud que hay entre los modelos de *spin-glass* en física con los algoritmos evolutivos. El trabajo del autor siempre está enfocado en el dominio de la química prebiótica, tratando de mostrar como surgen espontáneamente reacciones químicas autocatalíticas. Habla también de caos, atractores, cuencas de atracción y su trabajo principal, que son las redes booleanas estocásticas. Muestra como estas redes pueden aprender, utilizando técnicas evolutivas. Compara lo anterior con las redes neuronales. Habla del famoso *software Tierra*, de Tom Ray y las gramáticas aleatorias. Nos recuerda que Turing realizó un estudio que casi se pierde, sobre generación espontánea de patrones espaciales (manchas en la piel de los animales). De la página 407 en adelante habla sobre biología molecular de una manera muy técnica, por lo que no merece la pena leerse (a los biólogos les puede interesar, pero a los informáticos no).

- Ian Stewart (1998). *El segundo secreto de la vida*. Barcelona: Editorial Crítica.

Muestra el programa “Ameba”, una sopa de código aleatorio con mutaciones. Habla del poco conocido trabajo de Turing de computación descentralizada, para simular la creación de manchas en la piel de los animales. También de autómatas celulares, autoduplicación de Von Neumann, complejidad, evolución, emergencia, formas matemáticas de las plantas.

- Christoph Adami (1998). *Introduction to Artificial Life*. New York: Springer-Verlag.

Es un libro fácil de leer y toca casi todos los temas que vemos aquí. Habla sobre complejidad, entropía, información, información mutua, autocopía, virus... Explica muchos sistemas de vida artificial (Tierra, Avida, Amoeba, creatures de Karl Sims, autómatas celulares de von Neuman, Wolfram y Langton...). Habla de computabilidad de Turing, información algorítmica de Chaitin, caos y algo sobre biología.

- Carlos Eduardo Maldonado *et ál.* (2007). *Complejidad: ciencia, pensamiento y aplicación*. Bogotá: Universidad Externado de Colombia.

Escrito por varios investigadores, algunos de muy alta calidad. Hay artículos muy interesantes sobre caos en las redes de comunicaciones, medidas de complejidad, fractales, multifractales, un nuevo algoritmo para modelar sistemas fractal-multifractal, Prigogine, Gödel... Pero también deriva en cosas esotéricas y trabajos en temas que nada tienen que ver con la complejidad.

- Ilya Prigogine e Isabelle Stengers (1990). *Entre el tiempo y la eternidad*. Madrid: Alianza Editorial.

El premio nobel de química Ilya Prigogine ha escrito muchos libros divulgativos (*La estructura de lo complejo*, *¿Tan solo una ilusión?*, etc.), pero todos hablan de lo mismo con ligeras variaciones. Prigogine ha trabajado en el tema de sistemas complejos abordándolo desde conceptos físicos básicos (termodinámica de sistemas en desequilibrio). Muestra muchos ejemplos de sistemas caóticos (incluyendo la famosa reacción de Belousov-Zhabotinsky) y considera que todo el universo es evolutivo.

- Heinz R. Pagels (1991). *Los sueños de la razón*. Barcelona: Gedisa.

Explica diversas medidas de complejidad, fractales, el número de Feigenbaum, caos y aleatoriedad. Considera también los aspectos filosóficos y sociales.

- W. Brian Arthur (1990). Positive Feedbacks in the Economy. *Scientific American*, pp. 92-99, February 1990.

Es el primer economista que no cree en las ecuaciones de la economía clásica. Sus modelos no lineales predicen mejor el comportamiento de los mercados. Procede del Instituto de Santa Fé.

REFERENCIAS

Libros, artículos y enlaces web

- 20Qnet (2017). *20Q.net. The neural net on internet*. Recuperado el 2 de febrero de 2017. Disponible en: <https://goo.gl/pnoegz>
- Ashby, W. R. (1956). *An Introduction to Cybernetics*. Londres: Chapman and Hall.
- Atkins, P. W. (1992). *Cómo crear el mundo*. Barcelona: Editorial Grijalbo Mondadori.
- Barrow, J. D. (1999). *Imposibilidad. Los límites de la ciencia y la ciencia de los límites*. Barcelona: Gedisa Editorial.
- Bostrom, N. (2003). Are You Living in a Computer Simulation? *Philosophical Quarterly*, 53(211), pp. 243-255.
- Chaitin, G. J. (1992). *Information-Theoretic Incompleteness*. Singapore: World Scientific.
- _____. (1997). *Algorithmic Information Theory*. USA: Cambridge University Press.
- _____. (1997). *Information, Randomness & Incompleteness*. Papers on Algorithmic Information Theory. Singapore: World Scientific.
- _____. (1998). *The limits of mathematics*. Singapore: Springer-Verlag.
- _____. (1999). *The Unknowable*. Singapore: Springer-Verlag.
- _____. (2001). *Exploring Randomness*. London: Springer-Verlag.
- Clarke, A. C. (1979). «Marque F de Frankenstein». En *El viento del sol: relatos de la era espacial*. Madrid: Alianza Editorial, pp. 93-101.
- Dawkins, R. (1993). *El Gen Egoísta*. Barcelona: Salvat.
- DeDeo, S. (2018). *Information Theory for Intelligent People*. Recuperado el 25 de junio de 2018. Disponible en: <http://tuvalu.santafe.edu/~simon/it.pdf>
- Dennett, D. C. (1999). *La peligrosa idea de Darwin*. Madrid: Círculo de Lectores.
- Eplex Group (2017). *The Novelty Search Users Page*. Recuperado el 10 de abril de 2017. Disponible en: <https://goo.gl/ikyToR>
- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Massachusetts: Addison-Wesley Publishing Company Inc.
- Goodwin, B. (1998). *Las manchas del leopardo*. Barcelona: Tusquets Editores.
- Hart, V. y Case, N. (2017). *Parable of Poligons. A playable post on the shape of the society*. Recuperado el 20 de abril de 2017. Disponible en: <https://goo.gl/wwgUcD>
- Hoffman, P. (2000). *El hombre que solo amaba los números*. Barcelona: Editorial Granica.

- Hofstadter, D. (1979). *Gödel, Escher, Bach: un eterno y grácil bucle*. Barcelona: Tusquets Editores.
- Mandelbrot, B. y Hudson, R. L. (2006). *Fractales y finanzas*. Barcelona: Tusquets editores.
- Margulis, L. y Sagan, D. (1998). *¿Qué es el sexo?* Barcelona: Tusquets Editores.
- Maturana, H. R. y Varela, F. J. (1973). *De máquinas y seres vivos*. Santiago de Chile: Editorial Universitaria.
- McKinnon W. B., Nimmo, F., Wong, T. *et ál.* (2016). Convection in a volatile nitrogen-ice-rich layer drives Pluto's geological vigour. *Nature*, 534(7605), pp. 82-85. DOI: <https://doi.org/10.1038/nature18289>
- Miller, P. (2010). *The Smart Swarm: How understanding flocks, schools, and colonies can make us better at communicating, decision making, and getting things done*. New Jersey: Avery Publishing Group, Inc.
- Mitchel, M. (1999). *An introduction to genetic algorithms*. Cambridge: The MIT Press.
- Pattee, H. H. (2012). *Evolving Self-Reference: Matter, Symbols and Semantic Closure. From LAWS, LANGUAGE and LIFE: Howard Pattee's classic papers on the physics of symbols with contemporary commentary*, pp. 211-226. Springer. DOI: https://doi.org/10.1007/978-94-007-5161-3_14
- Peña, B. (2002) *Inestabilidades de Turing en Sistemas de Reacción-Difusión*. [Tesis Doctoral]. Pamplona: Universidad de Navarra.
- Prigogine, I. (1997). *La estructura de lo complejo*. Madrid: Alianza Editorial.
- _____. (1997). *El fin de las certidumbres*. Madrid: Taurus.
- Prusinkiewicz, P. y Lindenmayer, A. (2004). *The Algorithmic Beauty of Plants*. New York: Springer-Verlag.
- Searle, J. (1997). *El misterio de la conciencia*. Barcelona: Paidós.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(4), pp. 623-656.
- Smaldino, P. E. y McElreath, R. (2017). The Natural Selection of Bad Science. *R. Soc. open sci*, 3, 160384. Royal Society Publishing. DOI: <http://dx.doi.org/10.1098/rsos.160384>
- Tsallis, C. (2000). *Entropic Nonextensivity: A Possible Measure of Complexity*. Centro Brasileiro de Pesquisas Físicas. Rio de Janeiro: Santa Fe Institute Working Papers.
- Watts, D. J. (2010). *Seis Graus de Separação*. São Paulo: Leopardo Editora.
- Wissner-Gross, A. D. y Freer, C. E. (2013). Causal Entropic Forces. *Physical Review Letters*, 110(168702). DOI: <https://doi.org/10.1103/PhysRevLett.110.168702>

Películas y videos

- Bostrom, N. (2015). *What happens when our computers get smater than we are*. Recuperado el 16 de marzo de 2017. Disponible en: <https://goo.gl/wdjavs>
- Cameron, J. (1984). *Terminator*. USA: Orion Pictures.
- Chlebek, M. (2007). *Emergence*. Recuperado el 14 de abril de 2017. Disponible en: <https://goo.gl/qbNGfb>
- Connosco, A. R. (2011). *Emergence: Patterns on the Edge of Chaos - Introduction by Tiago Marques*. Recuperado el 2 de septiembre de 2017. Disponible en: <https://goo.gl/qRxKLU>
- Guaicoloro (2008). *Inestabilidad de Bénard Marangoni*. Recuperado el 26 de agosto de 2017. Disponible en: <https://goo.gl/bn8XrY>
- Hanson, J. (2017). *Are You Smarter Than A Slime Mold?* Recuperado el 3 de marzo de 2017. Disponible en: <https://goo.gl/n3pHkp>
- jNode (2016). *Structure Formation*. Recuperado el 2 de septiembre de 2017. Disponible en: <https://goo.gl/phF6St>
- Krauss, L. (2009). *Un universo desde la nada*. AAI web. Recuperado el 30 de abril de 2017. Disponible en: <https://goo.gl/fWyqmU>
- León, H. C. (2010). *Rayleigh Benard Convection, using a LB method on Matlab*. Recuperado el 2 de septiembre de 2017. Disponible en: <https://goo.gl/22GX5j>
- manf1234 (2017). *Self Organization*. Recuperado el 2 de septiembre de 2017. Disponible en: <https://goo.gl/zvdmYD>
- Openlb (2013). *Rayleigh Benard Thermal Convection Long Version*. Recuperado el 26 de agosto de 2017. Disponible en: <https://goo.gl/Gb5rwP>
- PaPookie (2006). *Pattern Formation in Nature*. Recuperado el 2 de septiembre de 2017. Disponible en: <https://goo.gl/LpkGX9>
- Ratcliff, W. C. (2014). *Experimental evolution of multicellularity*. Georgia Tech. Recuperado el 3 de mayo de 2017. Disponible en: <https://goo.gl/vteBGx>
- Reynolds, C. (1987). *Original 1986. Boids simulation*. Recuperado el 6 de mayo de 2017. Disponible en: <https://goo.gl/WAfYDA>
- Schoenfeld, I. (2013). *Introduction to Complex Systems: Patterns in Nature*. Recuperado el 2 de septiembre de 2017. Disponible en: <https://goo.gl/W3hG2E>
- Vila, C. (2010). *Nature by Numbers*. Recuperado el 2 de septiembre de 2017. Disponible en: <https://goo.gl/5bmRp9>
- Villatoro, F. R. (2010). *Todo es entropía*. NAUKAS. Recuperado el 12 de octubre de 2017. Disponible en: <https://goo.gl/utuJr3>

Tesis y trabajos de grado en EVALAB

Céspedes, M. A. (2008). *Incorporación de un parámetro fractal al algoritmo genético en problemas NP*. Cali: Universidad del Valle.

Gómez, J. A. (2001). *Desarrollo de una Máquina de Estados Finitos (FSM) evolutiva mediante algoritmos genéticos*. Cali: Universidad del Valle.

Mesa, N. (2014). *Framework para simulación de sistemas emergentes*. Cali: Universidad del Valle.

LIBERTAD

La libertad es no sentir miedo,
HELGA ROCÍO CALVO RIVERA

Una de las primeras cosas que necesita un sistema para aumentar su complejidad es tener autonomía para hacerlo. Debe tener una mínima libertad de elección para poder adaptarse a los problemas de su entorno y organizarse internamente de acuerdo a ello. Porque si no la tiene, así sea en grado mínimo, si las decisiones que toma son automáticas, entonces no tenemos un sistema independiente sino que es una pieza más del engranaje del universo. Pero, ¿qué es la libertad?

Podemos decir que la libertad es tener un gran abanico de posibilidades y poder elegir una de ellas. Esta es una definición sencilla e ingenua, y enseñada la refinaremos más. Pero está claro que quien no tiene opciones tampoco tiene libertad. Y quien, aun teniendo miles de opciones disponibles elige siempre la misma tampoco parece disfrutar de libertad. Si examinamos bajo esa óptica nuestro día a día, hay que reconocer que tenemos poca libertad ya que ejecutamos la misma rutina diaria de levantarnos cuando suena el despertador, asearnos, desayunar (probablemente lo mismo todos los días) e ir al trabajo o al estudio, y así sucesivamente en toda la jornada. Sufrimos de muchas restricciones físicas (no podemos volar por nosotros mismos), químicas, biológicas (no podemos dejar de comer y beber durante mucho tiempo, y menos de respirar o dormir) y sociales (hay que ir a trabajar). Desde luego hay pequeños momentos de libertad cuando decidimos de qué color va a ser la camisa que nos pongamos, o dónde vamos a viajar en vacaciones, pero calculando el porcentaje de esos instantes en el total de 24

horas diarias y 365 días del año, parece que somos poco libres. En este sentido, la libertad podría medirse en bits y eso es lo que haremos enseguida.

Por si eso fuera poco, los filósofos nos recuerdan que quizás no seamos libres en absoluto y que cuando creemos elegir entre un helado de fresa o uno de chocolate, nuestros genes tienen mucho que ver con ello, y la otra opción nunca la tomaríamos. También nos dicen que la libertad puede ser solo una ilusión, ya que el universo en el que vivimos sigue leyes físicas inexorables que posiblemente sean deterministas. Puesto que las decisiones las toma nuestro cerebro, que está formado por átomos, si el estado del universo volviera a repetirse, lo que ocurra a partir de allí también se repetiría. Y ello no deja ningún lugar a la libertad de elección.

A esto se le pueden añadir más matices: la interpretación actual de la mecánica cuántica que gobierna todos los fenómenos de este mundo incluye un componente estocástico ineludible²⁷, que podría ser la vía que nos devuelva la libertad. Sin embargo, a muchos filósofos eso no les sirve porque piensan que esos fenómenos estocásticos están decidiendo por nosotros. Creo que ese argumento tan pesimista es equivocado, pues por la teoría de juegos sabemos que en muchos casos nos conviene aleatorizar entre varias jugadas, con el objetivo de que el oponente no pueda predecirnos. Para ello se suelen emplear monedas o dados lanzados al aire. Y si finalmente logro ganar el juego, quien se lleva la recompensa soy yo. Y usé unos dados que son míos. No hay ninguna diferencia en que los dados estén implementados en el *hardware* mecánico cuántico de mi red neuronal. Son mis dados, son mis neuronas, y gracias a ellos he tomado decisiones.

Además, en cuanto comenzamos a estudiar estos temas nos damos cuenta de que unas definiciones dependen de otras: si no somos libres, si somos solo un engranaje más del universo, entonces no tiene mucho sentido creer que seamos inteligentes, ni mucho menos conscientes. Y todo es difuso: no hay una buena definición de inteligencia ni de libertad. Ni de consciencia, que parece ser una culminación de la inteligencia, pero tampoco nadie sabe definirla bien. A la vez que todo parece interrelacionado: para medir o entender objetos de mucha complejidad se requiere mucha inteligencia. Y para tener mucha inteligencia se requiere mucha complejidad. Para ser consciente se requiere ser libre. En definitiva, estos conceptos muestran dependencias mutuas, están interrelacionados y son difusos.

27 Aunque hay otras interpretaciones, como la de David Bohm de variables ocultas, pero que no concuerda con los resultados experimentales. Y otra donde lo estocástico no es tal, sino caos determinista.

Y no solo es que los filósofos debatan si somos libres o si esa libertad es apenas una ilusión y nuestro comportamiento está completamente determinado. Es que ni siquiera sabemos si vivimos en un mundo real o en una simulación. Si viviéramos en una simulación el concepto de libertad se volvería más endeble, pues unos hipotéticos seres que viviesen por fuera del simulador podrían mirarnos, mirar nuestro estado interno y saber absolutamente todo sobre nosotros. Ni tampoco sabemos si pueden existir seres más inteligentes o más conscientes que nosotros, o si después de la consciencia puede aparecer algún otro nuevo fenómeno más sofisticado. Ello se puede ver como una carencia nuestra, o sea, que somos incapaces de definir y entender estos conceptos. Pero también se puede ver como algo más fundamental y ese es el enfoque que se seguirá en este libro: son conceptos relativos unos a otros. En el correspondiente capítulo veremos que la inteligencia es relativa: podemos decir que un ente es más inteligente que otro. Pero no es fácil dar una medida absoluta de inteligencia de un ente, sin referirlo a otro u otros.

Con la libertad pasa algo parecido aunque no es exactamente lo mismo. Cuanta más libertad tenga un ente respecto a otro, más capacidad tendrá de manipularlo. Por ejemplo, un humano puede atar una zanahoria delante de la cabeza de un burro con el objetivo de que camine sin parar atado a una noria para que dé vueltas y saque agua²⁸. Mientras que el burro no puede manipular al humano. Al menos no tan fácilmente, pues se sabe que los bebés humanos intentan manipular a sus padres para atraer más atención, cuidados y comida. Y algunos animales domésticos también lo hacen. Pero un adulto humano que haya leído libros de psicología evolutiva puede evitar estas manipulaciones.

Manipular significa tener capacidad de predicción de lo que el otro va a hacer en función de diversas circunstancias, y ofrecerle la entrada adecuada para obtener el comportamiento deseado. El ente manipulado no se da cuenta de ello y, por tanto, cree que ha usado su libertad para tomar una decisión que cumpla sus objetivos (comerse la zanahoria, en el caso del burro) mientras que realmente está trabajando para cumplir los objetivos del otro (sacar agua para el humano). En este sentido, tener libertad significa que nadie pueda predecir lo que vas a hacer (Wolfram, 2002, p. 751). Tu comportamiento puede ser incluso determinista, pero si nadie puede predecirte, entonces nadie puede manipuarte. En resumen, eres libre.

28 Al objeto que se coloca delante del burro (la zanahoria o lo que sea) en el latín de la Roma clásica se le llamaba "manípulo".

Esta es una definición positivista y funcionalista, de la libertad. Es una definición poco esencialista, poco filosófica, más científica e ingenieril. Por ejemplo, el físico teórico Lee Smolin (2013, p. 148) también indica que “libertad” significa que no se puede deducir el comportamiento de una persona (o de una partícula) a partir de su pasado.

Y cualquier otra elucubración al respecto no tiene ningún sentido, sobre todo porque no se puede verificar, no es falsable, ni tampoco se puede implementar computacionalmente.

Recuadro 2: Definición positivista de la libertad

Un ente es tanto más libre cuanto menos se pueda predecir su comportamiento.

La libertad es entonces relativa a otro, al que te mira, al que quisiera manipularte. Y es de grado, porque pueden predecirte un 15%, un 27% o un 95% de las veces. Por cierto que esto tiene mucha relación con la “ley de la variedad requerida” de Ashby (1956), que se podría formular hoy en estos términos: la complejidad de un sistema debe ser mayor o igual a la de su entorno, si quiere controlarlo, entenderlo o adaptarse a él.

Tenemos un sujeto y queremos saber si es libre. El método a seguir es relativamente sencillo: le enfrentamos varias veces a un proceso en el cual el sujeto puede realizar dos acciones para resolverlo (A y B). Si observamos que toma la acción A algunas veces y la acción B otras veces, entonces es un indicador de que tiene libertad. Si el número de veces que elige A es idéntico al de veces que elige B , entonces maximiza la cantidad de información de la pregunta “¿elegirás A ?” y, con ello, maximiza su libertad.

Pero primero hay que superar algunas objeciones. Si el proceso es de tipo supervivencia, es decir, la acción A le permite seguir vivo mientras que la acción B lo mata, es de esperar que tome únicamente la acción A muchas más veces que la B , dado que la evolución lo habrá entrenado así. Con un resultado semejante no podemos deducir que es libre, sino que la evolución cableó sus respuestas para optimizar su supervivencia. Conforme el proceso no sea tan drástico (por ejemplo, en vez de vivir/morir, el proceso le permite decidir obtener comida/no comida, obtener mucha/poca comida, o cosas similares) la probabilidad de tomar la decisión equivocada va a ir en aumento, también por razones evolutivas. Por último, si el proceso ofrece dos opciones ninguna de las cuales afecta en lo más mínimo la supervivencia del ente, es de esperar que no haya habido allí ninguna presión evolutiva. Si el individuo elige cada opción A y B con igual probabilidad entonces podemos decir que es libre.

De todos modos, esto hay que elaborarlo más. Si el individuo toma las decisiones en el orden *ABABABAB* tampoco podemos decir que es libre a pesar de que el 50% de las veces elige *A* y el otro 50% elige *B*, debido a que hay una periodicidad que permite predecir lo que va a hacer. Y entonces para saber si el individuo es libre hay que verificar que no haya repeticiones ni ningún tipo de correlación entre sus respuestas. Desde esta perspectiva vemos que la libertad es indistinguible de una variable estocástica y hay que someterla a las mismas pruebas con que verificamos un generador de números al azar.

Ya que la complejidad la estamos midiendo con la fórmula de la entropía, tiene sentido modelar la libertad por medio de la información mutua: a mayor información mutua, menor libertad. Recordemos como se define la información mutua. Partamos de dos variables discretas que pueden tomar un conjunto de valores:

$$\begin{aligned} X &\in \{x_1, x_2, \dots, x_n\} \\ Y &\in \{y_1, y_2, \dots, y_m\} \end{aligned} \quad \text{Ec. 6}$$

La información asociada a un evento x_i de una variable X y que ocurre con probabilidad $p(x_i)$ es:

$$I(x_i) = \log_2 \left(\frac{1}{p(x_i)} \right) \quad \text{Ec. 7}$$

Y ya habíamos visto en el capítulo anterior que la cantidad de información de toda la variable X , llamada entropía de X es:

$$H(X) = \sum_{i=1}^n p(x_i) \log_2 \left(\frac{1}{p(x_i)} \right) \quad \text{Ec. 8}$$

Entonces la información mutua entre un valor concreto de X y otro de Y es:

$$I(x_i, y_j) = \log_2 \frac{p(x_i, y_j)}{p(x_i)} \quad \text{Ec. 9}$$

Y la información mutua media entre las dos variables X y Y es:

$$I(X, Y) = \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 \left(\frac{p(x_i, y_j)}{p(x_i)p(y_j)} \right) \quad \text{Ec. 10}$$

Por ejemplo, si las dos variables fueran independientes entonces

$$p(x_i, y_j) = p(x_i)p(y_j) \quad \text{Ec. 11}$$

Por lo que la información mutua entre ellas valdrá cero.

Y, en el otro extremo, si X y Y son la misma variable ($X=Y$) entonces:

$$p(x_i \wedge y_j) = \begin{cases} 0 & \text{si } i \neq j \\ p(x_i) & \text{si } i = j \end{cases} \quad \text{Ec. 12}$$

Por lo que la información mutua $I(X,X)$ valdrá lo mismo que $H(X)$:

$$I(X, X) = \sum_{i=1}^n p(x_i) \log_2 \left(\frac{1}{p(x_i)} \right) = H(X) \quad \text{Ec. 13}$$

Para terminar, recordemos las definiciones de entropía conjunta y entropía condicionada entre dos variables:

$$H(X, Y) = \sum_{i=1}^n \sum_{j=1}^m p(x_i \wedge y_j) \log_2 \left(\frac{1}{p(x_i \wedge y_j)} \right) \quad \text{Ec. 14}$$

$$H(X|Y) = \sum_{j=1}^m p(y_j) \sum_{i=1}^n p(x_i|y_j) \log_2 \left(\frac{1}{p(x_i|y_j)} \right)$$

De donde se deduce que:

$$I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y)$$

Tal y como aparece en la figura 29.

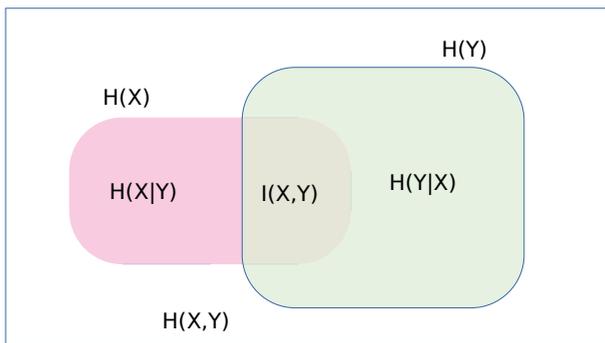


Figura 29. Relaciones de entropía de dos eventos

En la figura 30 vemos cómo es la interacción entre dos agentes sencillos X y Y . Cada uno tiene su propia complejidad, medida por la entropía $H(X)$ y $H(Y)$. La intersección de ambas es la información mutua $I(X, Y)$. Recordemos que es simétrica $I(X, Y) = I(Y, X)$. Entonces la libertad $L(X)$ y $L(Y)$ es la parte de la complejidad no compartida con el otro agente, es decir, lo que el otro agente desconoce y no puede predecir.

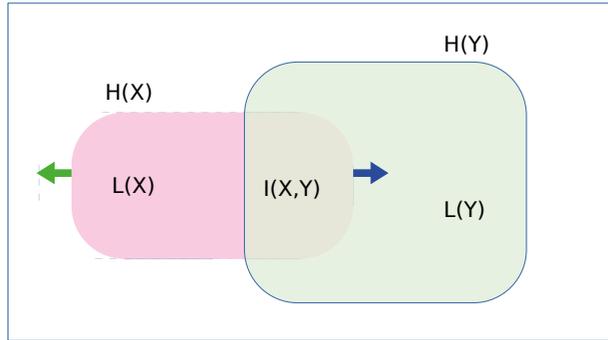


Figura 30. La complejidad de dos agentes sencillos

Siendo:

$$\begin{aligned} L(X) &= H(X) - I(X, Y) = H(X|Y) \\ L(Y) &= H(Y) - I(X, Y) = H(Y|X) \end{aligned} \quad \text{Ec. 15}$$

En la misma figura se indica cómo puede aumentar su complejidad el agente X : puede ser aumentando su libertad (flecha verde) o aumentando su inteligencia medida como capacidad de predicción del otro (flecha azul). O ambas, claro está. Y la situación es análoga para el agente Y .

De este modo, si tenemos un mundo con un conjunto de cuatro agentes $\{X, Y, Z, W\}$, la libertad de cada uno de ellos se puede medir como la intersección de su entropía con el complemento de la unión de la entropía de los demás:

$$L(X) = H(X) \cap (H(W) \cup H(Y) \cup H(Z))^c \quad \text{Ec. 16}$$

En la figura 31 puede verse un ejemplo, donde se indica la libertad de X como la zona rayada. Allí también se muestra la información mutua que tiene cada pareja de agentes.

Esta definición de libertad es fácil de entender y calcular, pero no es muy exacta. Únicamente valdría para agentes muy sencillos, de tipo reactivo, donde solo importan sus salidas, es decir, que no tengan estados internos o al menos que no tengan introspección. Solo así se cumple que lo que uno sabe del otro es lo que el otro sabe de uno. Ya que solo hay un lugar donde expresar la información que cada agente tiene (su salida), entonces son visibles tanto los comportamientos propios como las predicciones del comportamiento del otro.

controladas estrictamente en un laboratorio, un animal hace lo que le da la gana. En general, cuando un agente deja de ser reactivo porque tiene memoria, su libertad aumenta exponencialmente respecto a la cantidad de bits de esa memoria. La memoria que permanece oculta a los demás aumenta exponencialmente la libertad de cada individuo, pues con un bit se tienen dos posibilidades de actuar, con dos bits, cuatro posibilidades, con tres bits, ocho, y así sucesivamente.

Los procesos estocásticos generan eventos sin causas, lo cual es funcionalmente equivalente a decir que generan eventos cuyas causas no son computables. O que sí lo son, pero el tiempo necesario para computarlas es mayor que el que tomó para que ocurriera el evento, es decir, son eventos que no se pueden anticipar en tiempo real. En este sentido, el universo puede ser determinista, pero los eventos que produce no se pueden anticipar. De modo que existe la posibilidad de ser libres. Al respecto Dennett (1992, p. 140) compara la suerte de ganar una lotería normal, donde primero se venden los billetes y luego se extrae el número ganador, con una lotería invertida, donde primero se extrae el número ganador, que se guarda oculto en una caja fuerte, y después se venden los billetes. Matemáticamente ambas loterías son equivalentes y nadie se debe sentir estafado en el segundo juego. Después Dennett da otro paso: el número ganador se generó y se guardó en la caja fuerte antes de empezar el universo. En este caso, nuestra historia es determinista, pero eso no nos debe hacer sentir incómodos, pues sigue siendo equivalente matemáticamente a los dos casos anteriores, y nadie tiene oportunidad de sacar partido de ello.

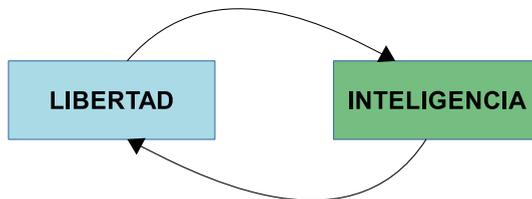


Figura 32. Bucle de realimentación positiva

En resumen, la libertad es evitar ser predicho y, como veremos más adelante, la inteligencia es predecir. Son dos conceptos complementarios y graduales. A la vez, el aumento de uno facilita el aumento del otro, por lo que conforman un bucle de realimentación positiva (Figura 32).

Desde luego que los niveles básicos de inteligencia no requieren libertad, pues la inteligencia puede verse como un proceso de búsqueda de óptimos y hay muchos algoritmos deterministas para ello. Pero la libertad sí se requiere cuando se desea lograr un nivel de inteligencia mayor.

Problema 4: Bar "El Farol"

En una pequeña ciudad hay un bar muy conocido al que todo el mundo desea ir el viernes en la noche. Lo malo es que el bar es pequeño, y si todo el mundo va, lo pasa mal porque no hay espacio. Concretamente, si menos del 60% de la población va al bar, entonces ellos lo pasan mucho mejor que si se hubieran quedado en casa. Por el contrario, si más del 60% de la población va al bar, entonces lo pasan peor que si se hubieran quedado en casa.

Cada una de las personas actúa independientemente (no se llaman por teléfono ni coordinan de ninguna manera). Y toman la decisión de ir o no ir de forma síncrona, en el mismo instante de tiempo.

¿Cuál es el mejor algoritmo que pueden usar para maximizar su disfrute?

¿CÓMO CONSEGUIR LIBERTAD?

La libertad se consigue siendo impredecible. Y el procedimiento matemático más rápido para ello es usar secuencias numéricas estocásticas, es decir, generadas al azar.

Lo malo del azar es que es totalmente impredecible, incluso para el sujeto que quiere gozar de libertad, con lo cual las acciones que genere a partir de ello pueden ser contraproducentes o incluso peligrosas para su propia supervivencia. Pero hay un fenómeno similar al azar aunque controlable: es el caos.

Los sistemas biológicos funcionan en el filo del caos, entre el orden y el desorden. Eso ha supuesto siempre un misterio, pero ahora es fácil de entender: la evolución u otros procesos han llevado a los sistemas biológicos al caos para lograr el máximo de libertad. Hay varios estudios del cerebro que muestran ese caos (Mejías, 2015). El médico neurofisiólogo colombiano Rodolfo Llinás nos contaba que el movimiento que notamos en la punta de los dedos cuando extendemos el brazo es el caos que se propaga por músculos y nervios desde el cerebro (todos tenemos ese temblor aunque no suframos de Parkinson), y que el movimiento más rápido que puede hacer el cuerpo es precisamente el de ese temblor. De modo que el caos nos permite realizar acciones a la más alta velocidad. Por otro lado, el caos es impredecible a largo plazo, pero predecible a corto plazo. Es incluso fácilmente controlable si conoces sus parámetros y los puedes variar. Es decir, el propio sujeto que genera el fenómeno caótico puede, en teoría, controlarlo. Pero los entes externos no lo tienen tan fácil. En ello radica el concepto de la libertad humana: mi caos lo controlo yo, y nadie más. Por supuesto luego viene la pregunta de quién es ese "yo", pero aquí no vamos a resolverla y lo dejaremos para el capítulo de la consciencia.

A su vez, como puedes ver en el libro anterior, este fenómeno se logra fácilmente incluso en una ecuación tan simple como:

$$f(x) = 4\lambda x (1-x) \quad \text{Ec. 17}$$

Esta ecuación es muy sencilla pues solo tiene una variable (x), dos constantes (4λ y 1), dos multiplicaciones y una resta. Sea como sea, tiene muy pocos bits de complejidad. Ni siquiera requiere computación completa para su cálculo. Por eso es que pensamos que la libertad es muy fácil de alcanzar a poco que aumente la complejidad de un sistema (Figura 33).

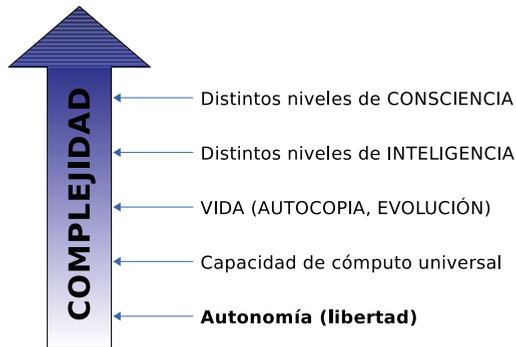


Figura 33. Con muy pocos bits de complejidad se consigue impredecibilidad, o sea, libertad

La libertad es como un banco: puedo ahorrar, es decir, acumular bits de libertad. Y posteriormente puedo gastarlos. Porque la libertad puede disminuir. Los bits de libertad que un sujeto tenga almacenados los va usando y se consumen conforme toma decisiones. Si se han agotado por completo, entonces para un observador externo será muy fácil predecir las acciones de ese sujeto. Por suerte, también se puede ganar libertad en el proceso de recibir información del entorno. Supongamos que tengo una cantidad pequeña, muy pequeña, pero significativa de libertad, concretamente 1 bit, y estoy entrando a una tienda futurista donde puedo comprar accesorios cerebrales como ampliaciones de memoria y cosas así. Me ofrecen un módulo que sirve para tomar decisiones libremente, con una capacidad de 3 bits. Usando mi bit de libertad, puedo decidir si lo compro o no lo compro. El bit de libertad con el que entro a la tienda lo voy a consumir tomando una decisión: si decido no comprar, me quedarán 0 bits de libertad, mientras que si decido comprar me quedarán $3-1=2$ bits de libertad. En este segundo caso habré aumentado mi libertad. Es importante entender que esa tienda futurista ya existe. Se llama colegio, universidad, biblioteca, Internet y cualquier otro sitio donde se puedan aprender cosas nuevas. Es decir, la libertad puede

amplificarse. Si se tiene algo de libertad, se puede conseguir más libertad todavía. Únicamente si no se tiene absolutamente nada de libertad es imposible amplificarla.

Los experimentos de Carnevale *et ál.* (2012) parecen indicar que así es. Allí se muestra también la subjetividad de la libertad: cuando es posible medir la aleatoriedad interna del sistema, entonces el sistema se vuelve predecible, es decir, carece de libertad. Podemos saber lo que la persona va a hacer antes de que ella misma lo sepa.

Por supuesto, también existe la posibilidad de usar la libertad para reducirla. Así ocurre en las adicciones al alcohol o las drogas y en cualquier proceso donde se dañe el propio cuerpo o la mente o se disminuya la futura libertad de acción.

Para terminar, un poco más de filosofía. Según Gribbin (2006, p. 113) hacen falta infinitas cifras decimales para calcular trayectorias caóticas. Para ello se requiere un computador infinitamente grande. Obviamente, el computador más grande que cabe dentro del universo es el propio universo. Eso significa que aunque el universo sea determinista, en la práctica ignora su propio futuro y eso es lo que genera libertad no solo en los seres humanos y los seres vivos, sino también en los procesos físicos considerados a largo plazo.

LIBERTAD Y EMERGENCIA

La libertad también puede darse debido a la emergencia. Si recordamos el capítulo anterior, cuando hay emergencia aparece una ontología, con nuevos objetos y nuevas reglas de interacción entre ellos. Pero no es exacta. Las reglas de bajo nivel son las que siguen funcionando. Las reglas de alto nivel son solo una forma conveniente, cómoda y corta de entender el sistema. Y entonces, a veces, las reglas de alto nivel fallan porque no capturan por completo lo que ocurre en el bajo nivel. Cuando eso ocurre decimos que hay un fenómeno estocástico superpuesto con el comportamiento de nuestro modelo de alto nivel. Pero no se trata de fenómenos sin causa, sino de causas de bajo nivel desconocidas por nosotros. En ese sentido, el sistema tiene libertad en el nivel superior, pues actúa siguiendo causas no modelables ni predecibles en ese nivel.

LIBERTAD Y TRANSICIONES DE FASE

Se tiene más libertad cuando mantenemos un sistema justo en la transición de fase entre dos estados estables. La razón de ello es que los estados estables son difíciles de modificar y fáciles de predecir. Seguramente hay realimenta-

ciones negativas que los mantienen en lo que son. Pero justo en la transición, cualquier pequeña perturbación lleva hacia una dirección u otra. El sistema es menos predecible y, por tanto, más libre.

Por eso debe ser que los sistemas, cuando alcanzan una cierta complejidad, tienen a buscar esas zonas. En muchos libros a este fenómeno se le llama *criticalidad autoorganizada* o “estar en el filo del caos”²⁹.

RESUMEN

En este capítulo proponemos una definición positiva de libertad como aquellos comportamientos que no pueden ser predichos por otros entes. Y hemos podido incluso formalizarla matemáticamente para el caso de seres simples, a través de la entropía y de la información mutua.

La libertad no consiste tanto en controlar lo que va a pasar, como en evitar que me controlen o manipulen.

En este sentido, matemáticamente hablando, el caos es un productor de libertad, ya que hay un horizonte de predicción más allá del cual no se puede saber que va a pasar. Por ello es un buen momento para repasar el capítulo del mismo nombre del libro anterior. Y la memoria que posea el individuo le permite almacenar la libertad, como si se tratase de un banco, para usarla después.

La libertad se puede amplificar por realimentación positiva. Si tienes un bit de libertad, puedes comprar dos bits de libertad. Aunque nada impide que lo malgastes y lo pierdas.

Además, hay bastante relación entre libertad e inteligencia, pues cada una permite el crecimiento de la otra.

PARA SABER MÁS

- Carlos Sánchez Fernández y Concepción Valdés Castro (2003). *Kolmogórov: el zar del azar*. Madrid: SL Nivola libros y ediciones.

No hay libros que traten la libertad desde una perspectiva computacional, pero hay muchos trabajos sobre entropía e información mutua. Este es uno de ellos, que recomiendo especialmente pues se relaciona con otros temas que veremos aquí.

²⁹ En muchos libros se dice “*at the edge of chaos*” y se traduce por “estar al borde del caos”, o sea casi en caos. Pero eso no es correcto. La traducción correcta sería “estar dentro del caos, que es una zona muy angosta, afilada, al borde entre el orden y el desorden”.

- Christoph Adami (1998). *Introduction to artificial life*. New York: Springer-Verlag.

Este libro no solo explica la mayoría de los temas que tocamos aquí y de las implementaciones en *software*, sino que también dedica tiempo a los aspectos teóricos de la entropía, la información y la complejidad.

- Gregory J. Chaitin:
- (1997). *Algorithmic Information Theory*. USA: Cambridge University Press.
- (1997). *Information, Randomness & Incompleteness. Papers on Algorithmic Information Theory*. Singapore: World Scientific.
- (1992). *Information-Theoretic Incompleteness*. Singapore: World Scientific.
- (2000). *Transcripción de la Conferencia Distinguida en la Escuela Distinguida de Ciencias de la Computación, Universidad Carnegie Melon*. Traducción de Andrés Becerra Sandoval.
- (2002). Información y azar. *Boletín de la Asociación Matemática Venezolana*, IX(1), pp. 55-81.
- (2002). Meta-mathematics and the foundations of mathematics. *EATCS Bulletin*, 77, pp. 167-179.
- (2002). *On the intelligibility of the universe and the notions of simplicity complexity and irreducibility*. arXiv:math/0210035 [math.HO]
- (2002). Paradoxes on randomness. *Complexity*, 7(5), pp. 14-21.
- (1999). *The Unknowable*. Singapore: Springer-Verlag.
- Verónica Becher y Gregory J. Chaitin. (2002). Another example of higher order randomness. *Fundamenta Informaticae*, 51(4), pp. 325-338. IOS Press.

Los anteriores libros y artículos de Chaitin (al que se considera el heredero intelectual de Gödel) hablan todos de lo mismo: la teoría algorítmica de la información como un límite computacional fundamental. La aleatoriedad forma parte intrínseca de las matemáticas, lo que ocurre es que a los matemáticos no les gusta explorarla. Todos estos libros son sencillos solo si te gustan los formalismos matemáticos.

REFERENCIAS**Libros, artículos y enlaces web**

- Carnevale, F., de Lafuente, V., Romo, R. y Parga, N. (2012). Internal signal correlates neural populations and biases perceptual decision reports. *PNAS*, 109(46), pp. 18938-18943. DOI: www.pnas.org/cgi/doi/10.1073/pnas.1216799109
- Dennett, D. (1992). *La libertad de acción*. Barcelona: Gedisa.
- _____. (2004). *Freedom Evolves*. New York: Penguin Books.
- Gribbin, J. (2006). *Así de simple*. Barcelona: Colección Drakontos, Editorial Crítica.
- Mejías, J. (2015). *Making room for a little chaos in your brain*. Recuperado el 3 de septiembre de 2017. Disponible en: <https://goo.gl/JmRNXJ>
- Smolin, L. (2013). *Time Reborn: from the Crisis in Physics to the Future of the Universe*. Houston: Houghton Mifflin Harcourt.
- Stewart, I. (1998). *El segundo secreto de la vida*. Barcelona: Colección Drakontos, Editorial Crítica.
- Wolfram, S. (2002). *A New Kind of Science*. Canadá: Wolfram Media Inc.

Películas y videos

- Dormael, J. V. (2009). *Mr Nobody*. Bélgica, Francia, Alemania, Canada: Pan-Europeenne.

COMPUTACIÓN COMPLETA

Wagner's music is better than it sounds,
Truth is the most valuable thing we have. Let us
economize it,

Giving up smoking is the easiest thing in the world.
I know because I've done it thousands of times,

The report of my death was an exaggeration,

MARK TWAIN

No one goes there nowadays, it's too crowded,

YOGUI BERRA

Las citas anteriores son no computables, porque contienen contradicciones internas en mayor o menor grado. Son autorreferencias que se niegan a sí mismas y enseguida hablaremos más sobre este tema. Sin embargo, nos hacen reír. Es más, el acto de entenderlas y reírse es un buen indicador de inteligencia.

Entonces ¿qué tiene la computación completa que es tan buena y tan mala a la vez, para lograr inteligencia? Lo veremos en el capítulo del mismo nombre. En este nos limitaremos a entender qué significa computación completa, a la manera de Turing. Comenzaremos por las estructuras computacionales más básicas, los grafos, y usando un tipo particular explicaremos que la computación universal es muy sencilla de lograr y que la propiedad que mejor la caracteriza es que puede simular cualquier otra máquina de cómputo universal. Terminaremos con las limitaciones que la propia computación de Turing se autoimpone. Sí. Has oído bien. Para lograr exactitud en el cómputo inevitablemente emergen tres limitaciones fundamentales respecto a lo que se puede llegar a computar.

GRAFOS

Los grafos son una estructura matemática simple que sirve para modelar casi cualquier cosa en sistemas discretos (en sistemas continuos es de poca utilidad). Un grafo tiene nodos y arcos que unen nodos, y puede ser orientado o no orientado, en función de cómo sean los arcos (Figura 34). Los grafos como rama de las matemáticas nacieron con el problema de los puentes de Königsberg, resuelto por Euler en 1736. En los últimos años esta rama ha recibido un nuevo impulso debido a las múltiples aplicaciones para modelar y entender redes de muy diversa índole, incluyendo los contactos en las redes sociales virtuales (Twitter, Facebook, G+, etc.), las redes de citaciones de artículos en revistas, las conexiones por avión entre aeropuertos, las redes de energía eléctrica, de agua, de gas o de teléfono, las redes ecológicas con predadores y presas, las redes de conexiones entre virus y sus proteínas objetivo, entre otras. Desde entonces ha recibido el nuevo nombre de *Network Science* (Ciencia de Redes) y si estás interesado en profundizar en este tema recomiendo leer los trabajos fundacionales de Mark Newman, Albert-László Barabási y Duncan J. Watts, que a su vez se basaron principalmente en las investigaciones de Erdős y Rényi de 1960 sobre grafos aleatorios. La aleatoriedad tiene una razón muy importante de ser a lo largo de todo este libro, y aquí la vemos de nuevo: la emergencia de propiedades nuevas en sistemas complejos debe ocurrir sin que haya una estructura *ad hoc* subyacente altamente improbable, pues de otro modo caeríamos en la mal llamada “filosofía del diseño inteligente”³⁰. Dicho con otras palabras: si hay fenómenos que se dan con certeza o con alta probabilidad a partir de condiciones iniciales aleatorias, entonces estamos en el camino correcto para entender la emergencia de la complejidad.

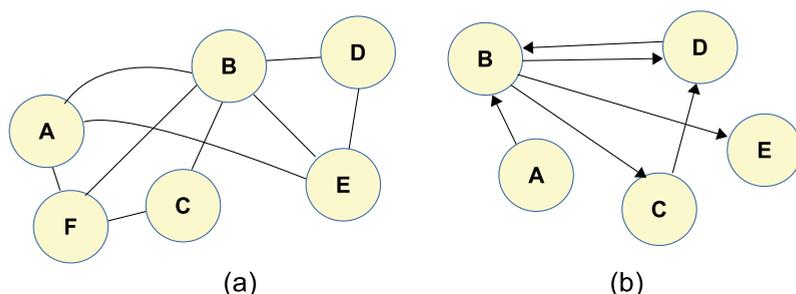
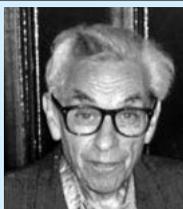


Figura 34. (a) Grafo no orientado con 6 nodos y 9 arcos; (b) Grafo orientado con 5 nodos y 6 arcos

30 Pues realmente no es una filosofía sino la base de la religión. Ver una crítica mordaz en Mackey (2008).

Algunas definiciones: los arcos pueden ser orientados (dirigidos) o no orientados (bidireccionales). El grado de un nodo es la cantidad de arcos que llegan o salen de él. Se puede definir el grado promedio de todo el grafo. La longitud de camino entre dos nodos es el mínimo número de arcos que hay que atravesar para llegar de uno al otro, y de ahí se puede calcular la longitud de camino promedio de todo el grafo. El diámetro de un grafo es la longitud de camino más grande que haya entre dos nodos cualesquiera. Si se considera la longitud de camino de un nodo respecto a todos los demás, su valor mayor es el radio de ese nodo; y entonces el centro del grafo es el nodo cuyo radio sea menor. Se puede asignar un grado de centralidad para cada nodo, en función de la diferencia entre su radio y el radio del centro del grafo. Hay otra forma de definir la centralidad de un nodo, que es la centralidad intermediaria y se calcula contando cuantos caminos más cortos entre cualquier par de nodos, pasan por el nodo en cuestión. El coeficiente de agrupamiento del grafo nos dice cuántos nodos forman subgrafos triangulares.



Fuente: Topsy Kretts (2003). Disponible en: <https://goo.gl/3QYV8C>

Personaje 3

Paul Erdős (1913-1996)

Paul Erdős fue un matemático húngaro desconocido para el común de la gente, a pesar de ser el más prolífico después de Euler. Representa muy bien al típico científico completamente absorto en su trabajo y con muy poca conexión con la vida cotidiana. Sus despistes son legendarios. Baste citar que no sabía atarse los zapatos y que no tenía casa propia, sino que viajaba por el mundo de congreso en congreso, hospedándose en casa de sus amigos matemáticos. Cuando uno de ellos (probablemente Stanislaw Ulam) fue hospitalizado a causa de un problema cerebral, Erdős se fue a acompañarlo poniéndole sin cesar problemas de ingenio, para asegurarse de la correcta recuperación del cerebro de su amigo. De él (o de alguno de sus amigos) procede el dicho de que "los matemáticos convierten el café en teoremas". También le gustaba dar pequeños premios en dinero, para quien fuera capaz de resolver algún desafío matemático. Para saber más de él puedes leer su biografía en un libro muy entretenido, *El hombre que solo amaba los números*, de Paul Hoffman.

Usando algoritmos sencillos de entender, aunque posiblemente de complejidad NP, se pueden medir las propiedades de estas redes. Podemos querer saber si hay mucha redundancia en las conexiones o si por el contrario

un fallo en un nodo puede aislar partes de la red. A lo mejor lo que queremos saber es si hay nodos con muy alta conectividad, llamados *hub*, desde los cuales es fácil acceder al resto de la red. A veces lo que nos interesa es más bien la influencia de los nodos sobre la red, por ejemplo, puede haber un nodo con pocas conexiones pero que son accesos directos a los *hub*. Y todo esto tiene interés para saber la velocidad de propagación de la información por la red: virus informáticos que saltan de un computador a otro por redes de contactos de *e-mail*, virus biológicos que se propagan por personas que están en contacto físico, rumores que se propagan por medio de canales electrónicos, incendios que se propagan de un árbol a sus vecinos, etc. El trabajo de investigación en esta área es muy entretenido y hay mucho por hacer y descubrir. Por ejemplo, en el 2014 María Andrea Cruz desarrolló con mucho éxito su trabajo de grado en EVALAB tratando de identificar agrupamientos de nodos en redes, usando técnicas de computación evolutiva y permitiendo solapamiento. En aquella época, Erdős seguramente no lo sabía, pero sus investigaciones sobre grafos iban a servir hoy día para dos cosas: analizar redes sociales virtuales, inexistentes en su época; y entender una forma de emergencia, cuando todavía no se hablaba de ese concepto. Para que te formes una idea de lo último, te propongo que intentes resolver el problema 5.



Problema 5: Teorema de la amistad

Fuente: Freepik. Disponible en: www.flaticon.com

En una fiesta hay N personas. ¿Cuál es el menor valor de N para que existan obligatoriamente al menos 3 personas que, o bien no se conozcan entre sí, o bien sí se conozcan entre sí?

El problema anterior es un ejemplo de emergencia determinista. A partir de cierto valor de N , aparecen propiedades nuevas que no existían antes. En otros casos, la emergencia se da de forma probabilista, es decir, no hay garantías sino solo una cierta probabilidad de que ocurra. La mayoría de las emergencias que vamos a ver en este libro son probabilistas, por lo que aquí dejo, en el problema 6, otro ejemplo determinista.

No es el objetivo de este libro entrar a detallar las redes complejas y sus aplicaciones³¹, pero sí indicar que, dependiendo del algoritmo empleado en

31 Ver, por ejemplo, Watts (2010).

su fabricación (aleatoria, de mundo pequeño, con conexión preferencial u otras), la red va a tener unos parámetros macroscópicos muy característicos (longitud del camino promedio, diámetro, coeficiente de agrupamiento, centralidad e influencia de cada nodo). Y ello permite hacer razonamientos por analogía entre redes similares aunque provengan de ámbitos muy distintos (biología, ciencias sociales, física, y otras áreas).

David G. Green y David Newth (2005) nos cuentan que los patrones de dependencias en modelos de matrices, sistemas dinámicos, autómatas celulares, semigrupos y conjuntos parcialmente ordenados son todos isomórficos con los grafos dirigidos. Y que el espacio de estados de cualquier autómata o conjunto de autómatas forma un grafo dirigido. Esto quiere decir que cualquier sistema complejo hereda sus propiedades de las de los grafos dirigidos. Además, hay una fuerte relación entre conectividad en un grafo y umbrales críticos para que emerja algo nuevo. O sea, se pueden observar cambios de fase conforme se añaden arcos de forma más o menos aleatorios a un grafo.



Problema 6: Galletas para todos

Fuente: dibujo realizado y cedido amablemente por Helga R. Calvo R. (2017)

Una fábrica de galletas dispone de cajas de empaque para 6, 9 y 20 unidades. ¿A partir de qué número N de galletas es posible atender cualquier pedido?

Por ejemplo, N no es 10 porque no hay forma de empaquetar 10 galletas en esas cajas, ni tampoco 11. Tampoco N vale 12, porque aunque 12 se pueden empaquetar en 2 cajas de 6, no es posible empaquetar 13 galletas.

Algunos de mis estudiantes acertaron con la solución, pero sin demostrarla. La estrategia fue escribir un *software* que iterase desde 6 hasta 1 000 000 de galletas probando a empaquetar esas galletas usando todas las combinaciones de empaque. Es un *software* muy sencillo de escribir. Al final encontraron un valor de N a partir del cual todos los pedidos podían ser empaquetados, pero... hasta 1 000 000 de galletas. Sin embargo, lo que exige el enunciado es probarlo hasta infinitas galletas. Eso no se puede hacer con esta estrategia. La programación convencional no abarca la noción de infinito.

¿Tienes tú alguna estrategia mejor?

En resumen, con los grafos se puede modelar prácticamente cualquier cosa digital, y en particular capturan bien las nociones de complejidad y de emergencia de propiedades.

COMPUTACIÓN UNIVERSAL

Un ejemplo de grafo muy particular son los autómatas celulares, donde las conexiones entre nodos siguen un patrón muy regular en forma de cuadrícula. Si es un autómata de dos dimensiones como el de la figura 35, con cada nodo conectado a sus vecinos de arriba, abajo, derecha e izquierda. O en forma de fila, si es de una dimensión. Todos los detalles se explican en el capítulo sobre autómatas celulares del libro anterior.

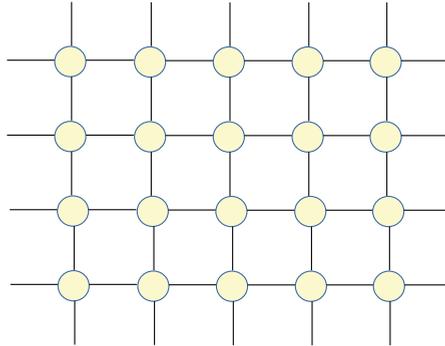


Figura 35. Un autómata celular 2D es un grafo muy simple y repetitivo

Los autómatas celulares son muy simples y eficientes computacionalmente hablando y permiten diseñar mundos artificiales con una noción de espacio básica, soportada en que para mover información de un nodo a otro hay que atravesar los nodos intermedios. No es posible dar saltos, todo el cómputo ocurre simultáneamente en cada nodo y no existe un control central.

Los autómatas celulares se han usado también para modelar interacciones entre partículas (Mitchell y Crutchfield, 1995) con éxito limitado, debido principalmente a que las métricas de distancias que surgen son de tipo Manhattan, muy alejadas de las *euclídeas* del mundo en que vivimos. Por ello, en EVALAB en el año 2013 demostramos que puede emerger una geometría *euclídea* discreta macroscópica usando grafos aleatorios a nivel microscópico, en el trabajo de grado Pablo Andrés Vélez.

Volviendo al tema que nos ocupa en este capítulo, la computación, en el 2002 Wolfram demostró que de los 256 tipos de autómatas celulares que existen en una dimensión, cuatro de ellos tienen capacidad de cómputo universal. Los detalles se pueden encontrar en el capítulo sobre autómatas celulares del libro anterior.

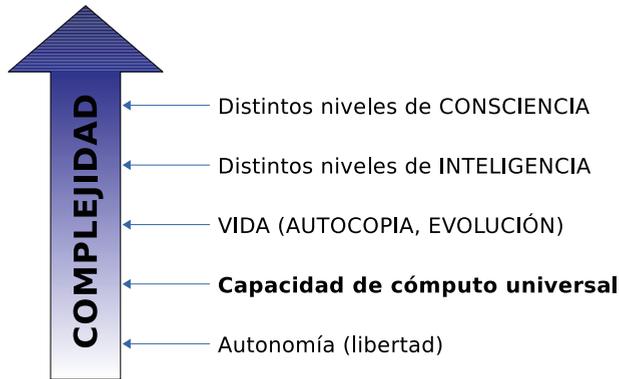


Figura 36. Con una complejidad del orden de 8 bits emerge la computación universal

Es decir, hay una alta probabilidad de lograr computación universal (un 2%) si por azar se logra una infraestructura similar a un autómata celular unidimensional, que requiere 8 bits de memoria y poco más. Concretamente, se requiere comunicación local entre objetos, es decir, cada objeto debe poder comunicar un bit de información al objeto de su derecha y al objeto de su izquierda. Esto es tan simple que ocurre en las células de un ser vivo, pero lo vemos también en las moléculas, los *spin-glasses*, los átomos, por ejemplo. Conclusión: en nuestro universo podemos contar con que aparezca espontáneamente la computación universal (Figura 36).

La capacidad de cómputo es una cosa, pero la universal es algo más potente, pues significa, entre otras cosas, que dentro de un computador así podemos simular otro, también con capacidad de cómputo universal. Y dentro de ese otro, otro más, y así sucesivamente, de manera indefinida.

Si queremos distraernos un poco de estos temas tan áridos, sin salir de ellos, hay muchas implementaciones de Máquinas de Turing extrañas, divertidas y sorprendentes. En IN6TV (2012) tenemos una construida físicamente en Lego, en Iker Garcia1996 (2016) hay otra dentro de *Minecraft* y en Bradbury (2012) podemos ver una máquina dentro de otra, implementadas en un autómata celular con las reglas de *LIFE* que vimos en el libro anterior.

¿QUÉ ES UN SIMULADOR?

El principio de Church-Turing-Deutsch (Deutsch, 1985) dice que una Máquina Cuántica Universal puede simular exactamente cualquier proceso físico. Entonces, según esta definición, el mundo es un computador cuántico. Ello no necesariamente implica que haya sido construido por alguien, ya que puede haber surgido espontáneamente a través de un proceso al azar, evolutivo u otros que desconocemos. Recordemos que es muy baja la

complejidad que se requiere para lograr una Máquina de Turing Universal Clásica. Podemos preguntarnos si hay una equivalencia entre Máquinas de Turing Universales Clásicas y Cuánticas, y la verdad es que todavía no se sabe. Durante el nacimiento de la computación cuántica se pensaba que iba a resolver problemas que no podía la clásica. Pero han pasado los años y todavía no hay un ejemplo de ello, por lo que comienza a pensarse que son equivalentes. En el momento en que ello se demuestre, también quedaría demostrado que nuestro mundo es computacional.

En cualquier caso, la tesis de este libro es que el mundo real es indistinguible de un mundo artificial que tenga similar complejidad y, por ello, es necesario explicar en detalle que un mundo artificial necesita un simulador para ejecutarse, así como las características que tiene, cómo funciona y cómo se construye.

Básicamente, un simulador es como un juego. Hay juegos en 2D, en 3D, con gafas de realidad virtual, con guantes hápticos y las tecnologías serán más y mejores en un futuro. Cuanto más complejo sea el mundo artificial, más sensación de realismo tendrá. Es más, el día que logremos conectar el juego directamente al cerebro, sin pasar por nuestros sentidos de la vista, oído y tacto podremos crear mundos con muchos más detalles, y más complejidad que el mundo real. Por ejemplo, podríamos hacer que los colores produzcan también sensaciones olfativas, y entonces el jugador no solo podrá ver el color amarillo si lo tiene delante, sino que también lo podrá oler si está cerca aunque fuera de su ángulo de visión. También podríamos crear nuevos sentidos para los ultrasonidos (como tienen los murciélagos), para los campos eléctricos (como tienen las anguilas) e incluso crear otros inexistentes e inimaginables en el mundo real. Regresar al mundo real será aburrido (y, ojo, que eso ya está ocurriendo con juegos muy adictivos).

Pero aquí hay algo que muchas personas no entienden bien. Un simulador va mucho más allá de eso. No solo permitirá a los humanos del mundo real sumergirse en el mundo artificial a la manera que han recreado en películas tan sugerentes como *Piso 13*, *Avalón*, *Matrix* y otras, sino que también el propio mundo simulado podrá tener seres artificiales propios, bien porque los hayamos diseñado ex profeso, bien porque hayan surgido allí a partir de las leyes de la física de ese mundo. Esto ya empieza a ocurrir en algunos juegos como “Spore”³² o “No Man’s Sky”³³ y solo es cuestión de tiempo que estos seres alcancen la complejidad necesaria para tener inteligencia y conciencia, y así poder reflexionar sobre el mundo en el que viven. Para estos seres,

32 Will Wright, *Electronics Arts*, 2008.

33 Hello Games, *Sony Interactive Entertainment*, 2016.

que no conocen nuestro mundo real, su mundo artificial será el real, pues no tienen ninguna experiencia que les lleve a añorar alguna característica externa. No pueden saber que hay un mundo más complejo que les abarca. E incluso si esta perspectiva parece muy fantástica, piensa que un niño real recién nacido al que se conecte su cerebro directamente a un juego (aunque sea poco sofisticado) aprenderá a vivir y a desarrollar su personalidad completamente sumergido en un mundo artificial. Si en este mundo no hay colores, sino solo grises, no los echará en falta. Si no hay sonido, tampoco. Etcétera. No puede imaginar cosas que jamás ha experimentado. Vivirá en su mundo artificial y no podrá saber que hay un mundo real más complejo ahí fuera. Esta es la razón principal que impide distinguir un mundo real de uno artificial: si estás dentro, no hay forma. Y siempre estás dentro.

Esto ya se puede hacer hoy día, si bien no se debe, por razones éticas. Hay que tener en cuenta que ya existen interfaces entre dispositivos electrónicos y las fibras nerviosas que permiten controlar directamente exoesqueletos con el pensamiento, de modo que es solo un problema técnico cablear todos los nervios que salen y entran del cerebro a un computador. Por otro lado, el ancho de banda sensorial humano es bastante más reducido de lo que se pensaba en un principio. Por ejemplo, la visión potencialmente debería requerir 10^{10} bit/s de información, que es lo que llega a la retina, pero lo cierto es que a la primera capa del córtex visual solo llega 10^4 bit/s (Raichle, 2010). Y para los otros sentidos se requiere menos ancho de banda. O sea, que los computadores actuales pueden recrear en tiempo real toda la información que recibimos.

El mundo que creemos real puede ser artificial y si uno hace las cuentas como Nick Bostrom, las probabilidades de que ello ocurra son altísimas porque hay una realimentación positiva (un proceso exponencial) involucrada.

Pero hay más cosas que tenemos que saber de un simulador. Una primera noción es que el simulador aísla dos mundos, el real y el simulado (Figura 37).

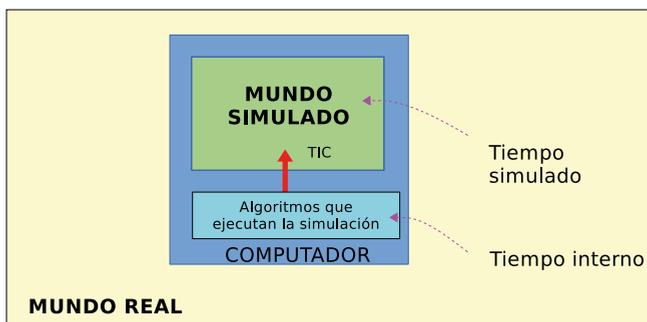


Figura 37. Dos mundos

Nosotros habitamos el mundo real y estamos sumergidos en sus “leyes de la Física”. Usándolas podemos construir el simulador. Pero dentro del simulador, el nuevo mundo no tiene por qué seguir esas leyes. Podemos crear otras nuevas hasta llegar a extremos inauditos. Examinemos unos cuantos ejemplos:

Nuestro espacio es de tres dimensiones y ello ya es de por sí bastante misterioso, y da para reflexionar, pues resulta que en un universo con una o dos dimensiones espaciales no es posible generar complejidad. Mientras que en cuatro o más dimensiones la complejidad es tanta que puede ser imposible de entender o manejar, impidiendo que existan estructuras autoorganizadas y adaptativas. O sea que tres dimensiones espaciales parece que nos lleva a la complejidad justa para que exista la vida. ¿Por qué nuestro mundo tiene tres dimensiones?

Pero en el mundo simulado el número de dimensiones espaciales puede ser distinto a tres. Si pensamos en un espacio absoluto, como un contenedor, basta que lo implementemos en *software* como una matriz. La matriz puede tener una dimensión (seguramente dará lugar a mundos muy aburridos), dos o tres (como en los juegos de tablero o de primera persona). Pero nada impide construir la matriz con cuatro dimensiones o con veintisiete o con cien mil millones. Seguramente saldrán mundos terriblemente complejos, pero son fáciles de construir en *software*. También se puede construir un mundo sin la noción del espacio (o sea, con cero dimensiones). Por ejemplo, cuando simulamos la evolución de poblaciones de predadores y presas con las ecuaciones de Lotka-Volterra, el espacio no tiene relevancia. Por así decir, cualquier predador puede comerse a cualquier presa porque todos se encuentran en el mismo punto espacial.

Además, las propiedades del espacio no tienen por qué ser iguales a las del nuestro. Por ejemplo, en el nuestro se cumple el teorema de Pitágoras (o sea, $h^2=a^2+b^2$, siendo h la hipotenusa y a, b los catetos de cualquier triángulo rectángulo). Mientras que el mundo virtual podemos implementarlo con autómatas celulares, usando la distancia Manhattan, de modo que se cumplirá en su lugar $h=a+b$. O decidir cualquier otra fórmula. O implementarlo con grafos aleatorios y que las distancias tengan ruido superpuesto o que cambien con el tiempo. O que las dimensiones no sean homogéneas (dos dimensiones pueden ser Euclídeas, otras dos Manhattan, otra permite moverse hacia un lado pero no hacia el otro, etc.). Como ya dijimos, en EVALAB (Vélez, 2013) diseñamos un mundo con grafos aleatorios evolutivos, de modo que aunque en distancias cortas la conexión entre nodos era al azar (lo cual podría simular saltos cuánticos como el efecto túnel) en distancias largas se cumple el teorema de Pitágoras.

Los objetos podrían tener masa inercial pero no masa gravitatoria, o al revés. O podríamos diseñarlo de modo que la energía no se conserve. O que no exista el concepto de energía. O que haya partículas distintas al electrón, quarks, fotones, etc. Y además podríamos cambiar el valor de cualquier constante de las que llamamos fundamentales, como la velocidad de la luz, la constante de Plank, la carga del electrón... O que no sea constante. O eliminar el concepto. O añadir conceptos nuevos. No hay límite a la imaginación.

O que no haya nada estocástico. Por ejemplo, cuando se lance una moneda al aire que siempre salga cara³⁴. O que siempre salga repetido tres veces seguidas cada resultado. O cualquier otra cosa. Cuando hacemos *software* tenemos libertad absoluta para cambiar el comportamiento de los objetos, y además es trivial de lograr.

El tiempo también lo podemos manipular pues en todo simulador hay dos ejes de tiempo: un tiempo del mundo simulado que podemos diseñar como queramos (puede tener incluso varias dimensiones); y un tiempo interno que coincide con el del mundo real externo, y que no es observable por hipotéticos seres que habiten dentro del simulador porque cuando corre el tiempo interno, el simulado (con el que laten los corazones de los seres virtuales) está detenido.

Para entender todo esto, vamos a contar *grosso modo* cómo funciona un simulador de circuitos digitales síncronos, usando como ejemplo el circuito de la figura 38, que tiene dos entradas conocidas (e_1, e_2), tres nodos internos (n_1, n_2, n_3) y una salida (s_1). El circuito está construido con una puerta *AND* de tres entradas, dos puertas *NOT* (que siempre son de una entrada) y un *flip-flop*.

34 El libro está dirigido a personas que sepan programar. Pero para las que no, es tan sencillo como definir:

```
def lanzar_moneda()
  return "CARA"
end
```

Mientras que en un mundo como el nuestro la función sería:

```
def lanzar_moneda()
  return rand < 0.5 ? "CARA" : "SELLO"
end
```

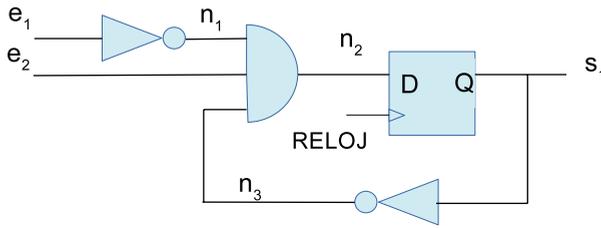


Figura 38. Un circuito digital sencillo

Para quien no recuerde cómo funcionan los circuitos digitales, cada señal tiene dos posibles niveles (0 y 1). Las tablas de verdad de las puertas están descritas en la figura 39. Mientras que el *flip-flop* es un bit de memoria que funciona así: la salida *Q* mantiene su estado permanentemente, y solo cambia cuando llega un pulso de *RELOJ*, copiando en ella lo que hubiera a la entrada *D*. Este circuito es muy sencillo, pero nos servirá para nuestros objetivos. Los circuitos reales tienen billones de puertas lógicas y *flip-flops*, y estos últimos comparten la misma señal de reloj (por eso se llaman síncronos). Con ello se logra que las salidas solo cambien en los instantes que llegan pulsos de *RELOJ*.

ENTRADAS	SALIDA	ENTRADA	SALIDA
000	0	0	1
001	0	1	0
010	0		
011	0		
100	0		
101	0		
110	0		
111	1		

(a)

(b)

Figura 39. Tabla de verdad de: (a) la puerta AND; (b) la puerta NOT

Este *RELOJ* es el reloj del circuito simulado (que coincide con el TIC de la figura 37). Es el que ve el usuario humano en la pantalla del computador, cada vez que avanza un pantallazo. También es el que percibirían unos hipotéticos seres que vivan dentro del simulador. Pero existe un tiempo interno que se desenvuelve entre los pulsos del reloj simulado, y que sirve para calcular los valores de los nodos internos, y que no ve el usuario humano desde su mundo externo. Tampoco lo podrían percibir unos hipotéticos seres dentro del simulador porque mientras se ejecuta el tiempo interno su vida está detenida, entre medias de dos tics de su reloj simulado. Por ejemplo, el simulador debe calcular el valor de n_1 negando la entrada e_1 , debe calcular

n_2 como un *AND* de n_1 , e_2 y n_3 , pero a su vez n_3 también hay que calcularlo como negación de s_1 . Y estos cálculos se hacen de una manera ingenua, en cualquier orden, de modo que pueden dar lugar a valores espurios que, con el paso del tiempo interno y más cálculo, finalmente se estabilizarán a los valores correctos. Lo ideal es que cuando todos los cálculos (realizados usando el tiempo interno) se hayan estabilizado, entonces llegue un pulso del *RELOJ* simulado que producirá las nuevas salidas observables al usuario humano (o un latido del corazón del hipotético habitante simulado).

Los cálculos internos en este caso son muy sencillos (ecuación 18), pero el orden en que se ejecuten estos tres pasos produce resultados distintos. Para el simulador es complicado deducir el orden correcto porque puede haber muchas dependencias cruzadas, por lo que habitualmente calcula estos pasos en cualquier orden y luego vuelve a repetir los cálculos hasta que todos los nodos internos llegan a un valor estable, que ya no cambia más. Allí termina la ejecución del tiempo interno y entonces el simulador envía un pulso por el *RELOJ* simulado, para que los cálculos efectuados produzcan cambios en las salidas.

$$\begin{aligned} n_2 &= \text{AND}(n_1, e_2, n_3) \\ n_3 &= \text{NOT}(s_1) \\ n_1 &= \text{NOT}(e_1) \end{aligned} \qquad \text{Ec. 18}$$

Pero no siempre es esto tan sencillo. Hay casos especiales donde los cálculos internos no convergen (ver fragmento de circuito en la figura 40, cuyos cálculos internos están en la ecuación 19). Aquí, si $n_1=0$ entonces $n_3=0$ y $n_2=1$. Y si volvemos a hacer los cálculos sale lo mismo. Perfecto. Ahora, si cambia el nodo $n_1=1$ entonces $n_3=1$ y $n_2=0$, pero si volvemos a calcular entonces $n_3=0$ y $n_2=1$. Y si otra vez lo calculamos sale $n_3=1$ y $n_2=0$, y así sucesivamente, es decir, los nodos n_3 y n_2 oscilan.

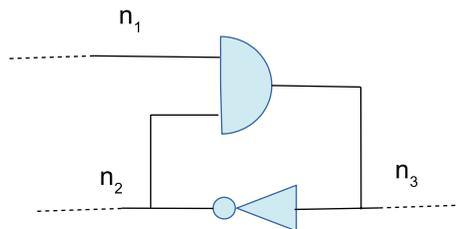


Figura 40. Circuito con problemas

$$\begin{aligned} n_3 &= \text{AND}(n_1, n_2) \\ n_2 &= \text{NOT}(n_3) \end{aligned} \qquad \text{Ec. 19}$$

Dependiendo de cómo esté diseñado el simulador de circuitos lógicos, puede que aborte la simulación indicando que hay un error (llamado “condición de carrera”) o puede que tenga un *timeout* que le obligue a detenerse, presentando como resultado cualquiera de los valores de la oscilación. Incluso si el simulador es sofisticado, puede aceptar ambos resultados y continuar la simulación mostrando al usuario la superposición de ambos. Al usuario le decimos “ambas posibilidades son válidas”, y ello traerá consecuencias sobre los resultados mostrados en los siguientes tics del *RELOJ* simulado que serán también superposición de otros resultados³⁵. Por ejemplo, si un nodo tiene una superposición de 0 y 1, y después viene una puerta *NOT*, su salida será una superposición de 1 y 0. Y si hacemos un *AND* entre las dos señales el resultado será siempre 0 desapareciendo la superposición. Es decir, en algunos casos la superposición puede aparecer y en otros desaparecer. La superposición puede ser molesta, pero no pasa nada si el resultado final del circuito no la tiene. La superposición de estados, que es un quebradero de cabeza para los que intentan comprender la mecánica cuántica, se entiende perfectamente suponiendo que el mundo sea computacional.

Y haciendo un pequeño paréntesis, estas oscilaciones son el equivalente a lo que los lógicos llaman “proposiciones indecidibles”, de las que habla el teorema de Gödel que veremos enseguida.

Con ello espero haber aclarado que en las simulaciones tenemos total libertad de crear (o no) espacio y tiempo de diversas dimensiones. Y de la misma manera podemos crear (o no) materia, energía, partículas y cualquier cosa que se nos ocurra. También es importante entender que las simulaciones que corren sobre computadores son discretas, es decir, no existe allí el concepto de continuidad. El espacio es discreto, el tiempo corre a saltos y cualquier cantidad elemental está cuantizada.

Otro concepto importante es que en un computador todo es determinista, no hay nada estocástico. Para conseguir impredecibilidad, lo único que se puede hacer es generar secuencias de números pseudoaleatorios utilizando fórmulas muy complicadas, pero debemos comprender que en el fondo son fórmulas fijas y deterministas. Otra alternativa es extraer información del mundo externo e inyectarla al mundo interno. En este caso, al mundo interno le llegan datos incorrelados con cualquier otra cosa que ya haya allí,

35 Recordemos que en nuestro universo existe esa posibilidad: si tenemos dos rendijas muy juntas y lanzamos hacia allí una partícula, ella puede pasar a la vez y simultáneamente por las dos rendijas. Este es el experimento más simple que nos muestra superposición en el mundo cuántico. Quizás haya llegado el momento de dejar de verlo como algo extraordinario, raro o mágico, y aceptar que es una consecuencia de que nuestro universo es computacional.

por lo que se pueden considerar estocásticos. Recordemos que, para ello, en Linux existe el flujo de datos */dev/random* que se puede leer por *software* y que contiene información sobre eventos externos (tiempos entre pulsaciones de teclas o entre movimientos del *mouse*, intervalos de llegada entre paquetes por el puerto de red y el ruido de la entrada del micrófono) convenientemente filtrados para eliminar autocorrelaciones.

Insisto en que los eventos externos no tienen por qué ser estocásticos en el ámbito del mundo real. Lo que importa es que sean incorrelados entre sí e incorrelados con los eventos del mundo interno simulado. Para unos hipotéticos seres que vivan dentro del simulador, un flujo de bits procedente del exterior y que esté incorrelado con cualquier otra información interna, lo interpretarán como estocástico.

LÍMITES COMPUTACIONALES

Hemos visto que la computación completa es fácil de lograr, pero la pregunta que vamos a analizar y tratar de contestar ahora es: ¿todo problema se puede resolver usando computación? Obviamente, hay muchos problemas del mundo real que no tienen solución, bien sea porque yo no controlo ninguna variable que pueda modificarlo, bien sea porque las restricciones son tantas y/o tan duras que ninguna solución puede satisfacerlas a todas. Sin embargo, hay otros problemas aparentemente sencillos, donde no vemos ningún obstáculo para encontrar su solución pero que en la práctica la computación no logra solucionarlos. ¿Por qué? ¿Cuáles son los límites que no puede franquear la computación? Veremos que existen tres limitaciones.

Este tema es importante no solo para la computación en sí, sino especialmente para la inteligencia artificial (y la natural) ya que los partidarios de la inteligencia artificial débil³⁶, principalmente Roger Penrose y John Searle, esgrimen estas limitaciones como una prueba de que la inteligencia artificial no puede lograr cualquier cosa, al contrario que la mente humana, a la que no vemos límites. Ellos no se dan cuenta que la mente humana padece de los mismos problemas, y que si somos creativos y logramos aparentemente superar limitaciones es porque empleamos el mismo truco que los algoritmos evolutivos o las redes neuronales artificiales: usamos aleatoriedad, no somos consistentes, avanzamos por el método de prueba y error, nos equivocamos con frecuencia y nuestros logros no se pueden atribuir a ningún ser humano

36 Los que piensan que los computadores pueden imitar algunos aspectos de la inteligencia humana, pero nunca podrán superarla, y que la mera imitación no equivale al fenómeno real de la inteligencia.

sino a toda la humanidad. Por eso, los algoritmos de inteligencia artificial deben renunciar a la exactitud y las garantías de convergencia, y abrazar el error, las contradicciones y el decir de vez en cuando a las preguntas planteadas “¡no lo sé!”, igual que hacemos los humanos.

Además, Penrose y Searle parecen desconocer que desde hace muchos años los algoritmos basados en la evolución, en razonamiento *bayesiano* y en redes neuronales artificiales han logrado éxitos impresionantes en campos que antes se creía que eran exclusivos de los humanos, como reconocimiento de rostros y de voz, conducción autónoma de vehículos o creación de obras de arte en música y pintura.

Por otro lado, para los partidarios de la inteligencia artificial fuerte³⁷ entre los que me encuentro, entender los límites de una técnica o herramienta es el primer paso para saltarlos. Dicho de otra forma, el 80% del camino que nos lleva a la solución de un problema consiste en entenderlo bien o, en otras palabras, en formular las preguntas correctas.

Entonces primero enumeraremos los límites conocidos a la computación, luego daremos un pequeño repaso a cada uno de ellos (pues realmente se conocen desde hace mucho tiempo, y existe literatura especializada en cada tema), y por último ofreceremos unas conclusiones sobre lo que ello implica para la inteligencia artificial.

Estos límites son el teorema de Gödel (que se traduce computacionalmente en el problema de la parada), la complejidad algorítmica y el teorema de *No-Free-Lunch*.

Teorema de Gödel

Para comenzar a entender a Gödel, pensemos primero en la importancia que tienen los sistemas axiomáticos (aplicación de reglas de manera mecánica). En la Grecia clásica, en vez de eso, existían unas reglas de razonamiento pero en lenguaje natural. Por ejemplo, el *Modus Ponens* de la figura 41:

Los hombres son mortales.
Sócrates es un hombre.
Luego Sócrates es mortal.

Figura 41. Modus Ponens

Sin embargo, esta forma de razonar está llena de problemas. Veamos un ejemplo en la figura 42:

³⁷ Los que creen que una simulación que capture los detalles esenciales de un proceso es indistinguible de ese mismo proceso.

Los chinos son numerosos.
Confucio es chino.
Luego Confucio es numeroso.

Figura 42. Modus Ponens equivocado

Otro ejemplo muy bonito de razonamiento defectuoso puede encontrarse en el cuento “Lo que le dijo la Tortuga a Aquiles” en el libro *Matemática Demente*, de Lewis Carroll, donde la Tortuga ingeniosamente obliga al ingenuo Aquiles a llevar a cabo una deducción, aparentemente simple, en infinitas proposiciones lógicas. Recordemos que en la Grecia clásica, la Tortuga y el gran corredor Aquiles ya se habían enfrentado en una carrera de 100 metros, donde Aquiles dio 50 metros de ventaja a la Tortuga. Para los griegos era una paradoja que Aquiles nunca alcanzase a la Tortuga, a pesar de ser 10 veces más veloz, pues cuando él hubiera recorrido los 50 metros que le separaban de ella, la Tortuga habría avanzado $50/10=5$ metros (la décima parte); y cuando Aquiles recorriera como un rayo esa distancia, la Tortuga ya no estaría allí, pues habría avanzado $5/10=0.5$ metros. El resultado para los filósofos griegos de aquella época era que Aquiles jamás alcanzaría a la Tortuga. Para nosotros, que ya conocemos el cálculo diferencial de Newton y Leibnitz, los griegos simplemente estaban calculando el instante del adelantamiento como límite de una serie infinita. Que resulte una serie infinita no quiere decir que el adelantamiento no vaya a ocurrir, sino que se eligió un tortuoso camino matemático para su cálculo. Y en el cuento de Carroll de nuevo se elige un procedimiento tortuoso para hacer una demostración lógica sencilla.

La forma que hemos encontrado de evitar estos defectos de razonamiento es formalizarlos, es decir, eliminar el lenguaje natural, el “sentido común” y la intuición, cambiándolos por un lenguaje formal matemático, esto es, un conjunto de verdades obvias de partida (axiomas) y un conjunto de reglas para encontrar nuevas verdades.

Así el primer razonamiento se traduce a lo siguiente: sea H el conjunto de los hombres, sea M el conjunto de los mortales y sea s , Sócrates. Entonces:

$$(\forall x: x \in H \rightarrow x \in M) \wedge (s \in H) \vdash (s \in M) \quad \text{Ec. 20}$$

Mientras que el segundo razonamiento se formaliza así: sea C el conjunto de los chinos, sea N el conjunto de los conjuntos que son numerosos y sea f , Confucio. Entonces:

$$(C \in N) \wedge (f \in C) \vdash \text{de aquí no se deduce nada} \quad \text{Ec. 21}$$



Fuente: The Life and Letters of Lewis Carroll (2010). Disponible en: <https://bit.ly/2ZRh2nj>

Personaje 4

Lewis Carroll (1832-1898)

Es el seudónimo del matemático Charles Lutwidge Dodgson, muy conocido por sus cuentos “Alicia en el país de las maravillas” y “A través del espejo”. Aprovechando que ya es de dominio público presentaremos otro cuento menos conocido, “Lo que la Tortuga le dijo a Aquiles”, escrito en 1894 y recopilado en el libro *Matemática Demente* (1995):

Aquiles había alcanzado a la tortuga y se había sentado cómodamente sobre su caparazón.

—¿De modo que ha llegado usted al final de nuestra carrera?— dijo la Tortuga. —¿Aun cuando consistía en una serie infinita de distancias? ¿Pensó que algún sabihondo había probado que la cuestión no podía ser realizada?

—Sí puede ser realizada— dijo Aquiles. —¡Ella ha sido realizada! Solivitur ambulando. Usted ve, las distancias fueron disminuyendo constantemente y así...

—Pero, ¿si hubieran ido aumentando?— interrumpió la tortuga. —Entonces ¿qué?

—Entonces yo no debería estar aquí— replicó modestamente Aquiles, —y a estas alturas usted hubiera dado ya varias vueltas al mundo.

—Me aclama, aplana, quiero decir— dijo la Tortuga —pues usted sí que es un peso pesado, ¡sin duda! Ahora bien, ¿le gustaría oír acerca de una carrera en la que la mayoría de la gente cree poder llegar con dos o tres pasos al final y que realmente consiste en un número infinito de distancias, cada una más larga que la distancia anterior?

—¡Me encantaría, de veras!— dijo el guerrero griego mientras sacaba de su casco (pocos guerreros griegos poseían bolsillos en aquellos días) una enorme libreta de apuntes y un lápiz. —¡Empiece, y hable lentamente, por favor! ¡La taquigrafía aún no ha sido inventada!

—¡El hermoso Primer Teorema de Euclides!— murmuró como en sueños la tortuga. —¿Admira usted a Euclides?

—¡Apasionadamente! ¡Al menos, tanto como uno puede admirar un tratado que no será publicado hasta dentro de algunos siglos más!

—Bien, en ese caso tomemos solo una pequeña parte del argumento de ese Primer Teorema: solo dos pasos y la conclusión extraída de ellos. Tenga la bondad de registrarlos en su libreta. Y, a fin de referirnos a ellos convenientemente, llamémoslos A, B y Z.

(A) Dos cosas que son iguales a una tercera son iguales entre sí.

(B) Los dos lados de este triángulo son iguales a un tercero.

(Z) Los dos lados de este triángulo son iguales entre sí.

Los lectores de Euclides admitirán, supongo, que Z se sigue lógicamente de A y B, de modo que quien acepte A y B como verdaderas debe aceptar Z como verdadera, ¿no?

—¡Sin duda! Hasta el más joven de los alumnos del Liceo— tan pronto como se inventen los Liceos, cosa que no sucederá hasta dentro de dos mil años —admitirán eso.

sigue

—Y si algún lector no ha aceptado A y B como verdaderas, supongo que aún podría aceptar la secuencia como válida.

—Sin duda que podría existir un lector así. Él podría decir “Acepto como verdadera la Proposición Hipotética de que si A y B son verdaderas, Z debe ser verdadera, pero no acepto A y B como verdaderas”. Un lector así procedería sabiamente abandonando a Euclides y dedicándose al fútbol.

—Y ¿no podría haber también algún lector que pudiera decir “Acepto A y B como verdaderas, pero no acepto la Hipotética”?

—Ciertamente podría haberlo. Él, también, mejor se hubiera dedicado al fútbol.

—¿Y ninguno de estos lectores— continuó la Tortuga —tiene hasta ahora alguna necesidad lógica de aceptar Z como verdadera?

—Así es— asintió Aquiles.

—Ahora bien, quiero que usted me considere a mí como un lector del segundo tipo y que me fuerce, lógicamente, a aceptar Z como verdadera.

—Una Tortuga jugando al fútbol sería... —comenzó Aquiles.

—Una anomalía, por supuesto— interrumpió airadamente la Tortuga. —¡No se desvíe del tema, Primero Z y después el fútbol!

—¿Debo forzarlo a aceptar Z, o no?— preguntó Aquiles pensativamente. —Y su posición actual es que acepta A y B pero NO acepta la Hipotética...

—Llámosla C— dijo la tortuga, —pero no acepto que:

(C) Si A y B son verdaderas, Z debe ser verdadera.

—Esa es mi posición actual— dijo la Tortuga.

—Entonces debo pedirle que acepte C.

—Lo haré así— dijo la Tortuga —tan pronto como lo haya registrado en su libreta de Apuntes. ¿Qué más tiene anotado?

—¡Solo unos pocos apuntes!— dijo Aquiles agitando nerviosamente las hojas. —¡Unos pocos apuntes de las batallas en las que me he distinguido!

—¡Veo que hay un montón de hojas en blanco!— observó jovialmente la Tortuga. —¡Las necesitamos todas!— Aquiles se estremeció. —Ahora escriba mientras dicto:

(A) Dos cosas que son iguales a una tercera son iguales entre sí.

(B) Los dos lados de este triángulo son iguales a un tercero.

(C) Si A y B son verdaderas, Z debe ser verdadera.

(Z) Los dos lados de este triángulo son iguales entre sí.

—Debería llamarla D, no Z— dijo Aquiles. —Viene después de las otras tres. Si acepta A y B y C, debe aceptar Z.

—¿Y por qué debo?

—Porque se desprende lógicamente de ellas. Si A y B y C son verdaderas, Z debe ser verdadera. No puede discutir eso, me imagino.

—Si A y B y C son verdaderas, Z debe ser verdadera— repitió pensativamente la Tortuga. —Esa es otra Hipótesis, ¿o no? Y, si no reconociera su veracidad, podría aceptar A y B y C, y todavía no aceptar Z, ¿o no?

—Podría— admitió el cándido héroe, —aunque tal obstinación sería ciertamente fenomenal. Sin embargo, el evento es posible. De modo que debo pedirle que admita una Hipótesis más.

—Muy bien, estoy ansioso por admitirla, tan pronto como la haya anotado. La llamaremos 'D'. Si A y B y C son verdaderas, Z debe ser verdadera. ¿Lo ha registrado en su libreta de apuntes?

—¡Lo he hecho!— exclamó gozosamente Aquiles, mientras guardaba el lápiz en su estuche. —¡Y por fin hemos llegado al final de esta carrera ideal! Ahora que ha aceptado A y B y C y D, por supuesto acepta Z.

—¿La acepto?— dijo la Tortuga inocentemente. —Dejémoslo completamente claro. Acepto A y B y C y D. Suponga que todavía me niego a aceptar Z.

—¡Entonces la Lógica le agarraría del cuello y le forzaría a hacerlo!— replicó triunfalmente Aquiles.

—La Lógica le diría, “¡No se puede librar. Ahora que ha aceptado A y B y C y D, debe aceptar Z!”. De modo que no tiene alternativa, usted ve.

—Cualquier cosa que la Lógica tenga a bien decirme merece ser anotada— dijo la Tortuga —de modo que regístrela en su libro, por favor. La llamaremos “E”. Si A y B y C y D son verdaderas, Z debe ser verdadera. Hasta que haya admitido eso, por supuesto no necesito admitir Z. De modo que es un paso completamente necesario, ¿ve usted?

—Ya veo— dijo Aquiles. Y había un toque de tristeza en su tono de voz.

Aquí el narrador, que tenía urgentes negocios en el banco, se vio obligado a dejar a la simpática pareja y no pasó por el lugar nuevamente hasta algunos meses después. Cuando lo hizo, Aquiles estaba aún sentado sobre el caparazón de la muy tolerante Tortuga y seguía escribiendo en su libreta de apuntes que parecía estar casi llena.

La Tortuga estaba diciendo —¿Ha anotado el último paso? Si no he perdido la cuenta, ese es el mil uno. Quedan varios millones más todavía. Y le importaría, como un favor personal, considerando el rompecabezas que este coloquio nuestro proveería los Lógicos del siglo XIX. ¿Le importaría adoptar un retruécano que mi prima la Tortugacuática Artificial hará entonces y permitirse ser renombrado “Aquiles el Sutiles”?

—¡Como guste!— replicó el cansado guerrero con un triste tono de desesperanza en su voz, mientras sepultaba la cara entre sus manos. —Siempre que usted, por su parte, adopte un retruécano que la Tortugacuática Artificial nunca hizo y se permita renombrarse “Tortuga Tortura”.

Podemos ver que usando lenguajes formales evitamos las ambigüedades de los lenguajes naturales. Los lenguajes formales son un gran invento. Veamos un poco cuál fue su origen.

En el siglo XVII nace la física moderna, de la mano de Galileo, así como el método científico: los sucesos que antes eran arbitrarios pueden ahora medirse de forma repetible y modelarse con ecuaciones matemáticas, las cuales se verificaban de forma tan precisa que se decía y pensaba que los objetos del mundo material estaban cumpliendo leyes, leyes de la física claro, pero leyes, como las que impone un rey a sus súbditos. Esto dio lugar a un paradigma, es decir, una forma de ver y entender el mundo, llamado mecanicismo: todos los objetos materiales siguen leyes, no hay nada arbitrario, todo se puede predecir.

El físico y matemático Laplace fue quien mejor entendió y formuló esta idea, llamada determinismo: conociendo las condiciones iniciales de un sistema se puede predecir cómo va a cambiar en el tiempo. Y si conociéramos todas las posiciones iniciales y velocidades de todas las partículas del universo, podríamos calcular sus trayectorias desde ese momento en adelante, de modo que nada de lo que ocurra en el universo nos tomaría por sorpresa,

al poder predecirlo por adelantado. Este paradigma alcanzó también la matemática: Descartes y Leibnitz pensaban que se podía construir un método mecánico universal (lo que hoy llamamos un algoritmo) para solucionar cualquier problema que pudiera ser formalizado.

El siglo XIX fue el apogeo del determinismo en la Física. Pensemos que prácticamente todos los fenómenos físicos conocidos en aquella época tenían una explicación por medio de una ley, es decir, una ecuación. Se creía que se conocía todo acerca del universo. Es en ese ambiente que Bertrand Russell y Alfred North Whitehead escriben los *Principia Mathematica* donde formalizan la lógica y la matemática de los números enteros, trabajando en el “programa de Hilbert”, que intentaba construir toda la matemática a partir de un conjunto finito de axiomas y demostrar que el resultado era coherente.

En 1931 Kurt Gödel demostró que los *Principia Mathematica* no podían ser a la vez completos y coherentes, echando por tierra el programa de Hilbert que pretendía mecanizar las matemáticas. Concretamente, el teorema de Gödel dice que, si tenemos cualquier sistema axiomático suficientemente complejo, ocurre una de estas dos cosas:

- Es incoherente: contiene contradicciones internas, es decir, contiene simultáneamente una proposición y su negación. Si un sistema es incoherente entonces no vale para nada, puesto que a partir de la contradicción es posible demostrar cualquier proposición falsa.
- Es incompleto: existen verdades que no son demostrables desde dentro del sistema. Y este sería el origen de uno de nuestros límites computacionales para la inteligencia artificial, porque viene a decir que existen cosas que son verdad, pero que no se puede demostrar que lo sean. Esto aparenta ser una grave limitación para la inteligencia artificial.

Por supuesto, una verdad indemostrable en un sistema se puede demostrar ampliando los axiomas (o las reglas de ese sistema) añadiendo a ellos esa nueva verdad. Pero eso crea nuevas verdades indemostrables en el sistema ampliado. De modo que la incompletitud es intrínseca a los sistemas formales.

Este teorema es aplicable no solo a los *Principia Mathematica* (matemáticas de números enteros y lógica) sino a cualquier sistema formal axiomático suficientemente complejo.

Y ¿qué significa “suficientemente complejo”? Pues las matemáticas de los números enteros con las operaciones de suma y multiplicación ya son suficientemente complejas. La lógica de primer orden también es suficientemente compleja. Hay sistemas más simples —como los números enteros

solo con la operación de suma, los números reales (sin la noción de entero) con las operaciones de suma y multiplicación y la geometría— en los que no se aplica el teorema de Gödel.



Fuente: Public Domain. Disponible en: <https://goo.gl/9A2pQd>

Personaje 5

Lewis Carroll (1832-1898)

El matemático y filósofo Kurt Gödel nació en lo que hoy se conoce como Brno en la República Checa y se vio obligado a emigrar a los Estados Unidos, como tantos otros científicos, por el miedo al nazismo. Quizás por ello su personalidad tenía bastantes peculiaridades, como la de otras personas que se dedican intensamente a la ciencia, en este caso a las matemáticas. Dos anécdotas ilustran sus excentricidades: al ir a pedir la nacionalidad estadounidense, se ofuscó mucho porque descubrió que la constitución de USA contenía una contradicción lógica. Por suerte lo acompañaban sus amigos, Albert Einstein y Oskar Morgenstern, que no le permitieron discutir eso con el juez que le iba a dar la nacionalidad.

La otra anécdota, más dramática, se refiere a que solo recibía la comida que le preparaba su mujer porque temía ser envenenado. Murió de hambre, una vez que a ella tuvieron que hospitalizarla. Esto puede parecer sorprendente, pero con tanta información que aparece hoy en día recomendando evitar productos peligrosos (fito-hormonas, sulfitos, azúcar, aceite de palma, gluten, por citar algunos), cualquiera de nosotros podría terminar de la misma forma.

Aun cuando no es muy conocido por el público en general, Gödel fue un personaje extraordinario que cambió por completo la matemática, a la que despojó de sus ropajes de perfección.

Un libro excelente donde puede encontrarse una demostración asequible de este teorema es *Gödel, Escher, Bach: un grácil e infinito bucle* de Douglas Hofstadter, donde también se tocan aspectos filosóficos y prácticos sobre computación, evolución, creatividad e inteligencia natural y artificial. En este libro, Hofstadter propone el sistema axiomático MIU, que dejamos como ejercicio en el problema 7.

Problema 7: Acertijo MU

Hay 3 símbolos en este sistema: {M, I, U}

Hay 1 axioma: MI

Hay 4 reglas de inferencia:

Regla 1: De αI se deriva αIU

Regla 2: De $M\alpha$ se deriva $M\alpha\alpha$

Regla 3: De $\alpha III\beta$ se deriva $\alpha U\beta$

Regla 4: De $\alpha UU\beta$ se deriva $\alpha\beta$

donde las variables α y β representan cualquier cadena de símbolos, incluida la cadena vacía.

La pregunta es: ¿es MU una verdad en este sistema?

Por ejemplo, aplicando la regla 2 al único axioma sale MII, que entonces pasa así a ser una nueva verdad del sistema. Y aplicando de nuevo la regla 2 a MII sale MIIII que pasa a ser una nueva verdad. Y aplicando la regla 3 a MIIII sale MUI y también MIU que pasan a ser dos nuevas verdades. Y si comenzamos aplicando la regla 1 a MI sale MIU que pasa a ser una nueva verdad.

Vamos a ver un esbozo informal de la demostración de Gödel, que no es muy difícil pero confunde un poco. Se pueden encontrar otras demostraciones más completas en Smullyan (1987), Nagel y Newman (1994) y Hofstadter (1979). Básicamente Gödel tuvo que dar tres pasos:

- Expresar todas las proposiciones del sistema axiomático usando los símbolos de la lógica de primer orden.
- Codificar las proposiciones en un número de identificación de una manera ingeniosa.
- Plantear una proposición que hable de sí misma (a esto se le llama diagonalización o, en este caso concreto, *gödelización*), utilizando la codificación anterior y generando una contradicción.

Para establecer el procedimiento a seguir primero se deben aceptar algunos supuestos sobre las operaciones básicas permitidas que deben ser suficientemente simples, que no den lugar a ambigüedad y que sean ejecutables en un tiempo finito por medio de un algoritmo. Veremos solo algunos detalles.

El alfabeto de símbolos que requiere la lógica de primer orden es:

$a x f p \neg \wedge \vee \rightarrow \leftrightarrow \forall \exists () , 0 1 2 3 4 5 6 7 8 9 10 11 12$ etc.

Donde las constantes lógicas se expresan con a seguido de un número $a_0 a_1 a_2 \dots$, las variables lógicas con x seguido de un número $x_0 x_1 x_2 \dots$ las funciones³⁸ son f seguido de un número $f_0 f_1 f_2 \dots$ y las proposiciones son p seguido de un número $p_0 p_1 p_2 \dots$. El resto son la negación, la conjunción, la

³⁸ Las funciones en nuestro caso son las operaciones de suma + y multiplicación * de números enteros, pero en un caso más general son cualquier tipo de función.

disyunción, la implicación, la doble implicación, el cuantificador universal, el cuantificador existencial, los paréntesis, la coma y los números naturales incluyendo el cero.

Hay muchas formas de codificar estos símbolos en números enteros³⁹ y una de ellas la podemos ver en la tabla 1, que usa los números impares del 3 en adelante.

Tabla 1. Numeración de Gödel

Símbolo	a	x	f	p	¬	∧	∨	→	↔	∀	∃	()	,	0	1	2	3	etc.	
Código	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35	37	etc.

Usando los números primos como base y estos códigos como exponente, y haciendo el producto de todo ello, podemos codificar cualquier proposición de manera biunívoca. Por ejemplo, siendo x_1 y x_2 números naturales y f_1 la relación *mayor que*, se puede escribir una proposición que diga que los números naturales son infinitos así:

$$\text{Proposición } p_1: \forall x_1 \exists x_2 f_1(x_1, x_2) \tag{Ec. 22}$$

$$3^{21} \cdot 5^5 \cdot 7^{33} \cdot 11^{23} \cdot 13^5 \cdot 17^{35} \cdot 19^7 \cdot 23^{33} \cdot 29^{25} \cdot 31^5 \cdot 37^{33} \cdot 41^5 \cdot 43^{35} \cdot 47^{27} = \text{NG1} \tag{Ec. 23}$$

Se reservará el número primo 2 para indicar que a continuación sigue una secuencia de proposiciones. La coma dentro de las funciones se puede ignorar sin generar ambigüedad. Las secuencias de proposiciones se van a ligar con la conjunción (\wedge).

Si hacemos las operaciones matemáticas de la ecuación 6 da un número entero enorme que no soy capaz de poner aquí, y que vamos a llamar el número de Gödel de la proposición p_1 (NG1). Pero lo cierto es que ese número solo admite una única descomposición en factores primos, cuyos exponentes nos devuelven los códigos de los símbolos empleados en la fórmula. Por ello, cada número entero se puede convertir en una única fórmula, de modo que esta codificación es biyectiva. Por supuesto que, a partir de un número arbitrario, su descomposición en factores primos puede generar fórmulas mal construidas, pero eso no importa. Lo que importa es que toda fórmula bien construida tiene un número de Gödel único.

Miremos otro ejemplo. La proposición

$$\exists x x > 2 \cdot x + 1 \tag{Ec. 24}$$

³⁹ Por ejemplo, Wiki (2017a, 2017b).

Tiene una única variable libre x . Y no podemos decir que esa proposición sea verdadera ni falsa. Pero para el valor concreto de $x=3$ se convierte en la sentencia $3 > 2 \cdot 3 + 1$ que es falsa. Mientras que para el valor $x=-2$ se convierte en la sentencia $-2 > 2 \cdot (-2) + 1$ que es verdadera.

Ahora imaginemos que tenemos una secuencia de proposiciones $p1$ $p2$ $p3$ $p4$ cuyos respectivos números de Gödel son NG1 NG2 NG3 NG4 con las cuales se logra demostrar otra proposición $p5$, cuyo número de Gödel es NG5. Entonces se puede calcular el número de Gödel de la secuencia $p1 \wedge p2 \wedge p3 \wedge p4$ así:

$$2 \cdot 3^{NG1} \cdot 5^{13} \cdot 7^{NG2} \cdot 11^{13} \cdot 13^{NG3} \cdot 17^{13} \cdot 19^{NG4} = NG6 \quad \text{Ec. 25}$$

Recordemos que el primo 2 lo ponemos porque ahora esto no codifica una proposición sino una secuencia de proposiciones. Y las conjunciones, cuyo código es 13, sirven para ligar esas proposiciones. De este modo, cuando descompongamos NG6 en factores primos y uno de ellos salga 2 sabemos que los otros factores son proposiciones que se deben descomponer a su vez. Por eso es que en los códigos de los símbolos empleamos solo los números impares, para evitar ambigüedad.

En general no va a dar la casualidad de que NG5 coincida con NG6 y, de hecho, lo normal es que NG6 sea muchísimo mayor a NG5.

En español podemos decir que de la secuencia de proposiciones codificada con número de Gödel NG6 sirve para demostrar la proposición con número de Gödel NG5. Definamos entonces la función demostración f_2 .

$$\exists x_1 f_2(x_1, x_2) \text{ significa que la proposición } x_2 \text{ es demostrable.} \quad \text{Ec. 26}$$

Dado que cada número de Gödel identifica biunívocamente una proposición, entonces podemos decir que es verdad que $f_2(NG6, NG5)$, para esos valores concretos.

Definamos también la función de sustitución:

$$f_3(x_1, x_2, x_3) \quad \text{Ec. 27}$$

Significa que x_3 es el número de Gödel que sale al sustituir por una constante x_2 todas las variables libres (no cuantificadas) de la fórmula cuyo número de Gödel es x_1 .

Definamos otra función de autosustitución, como la anterior, pero haciendo $x_1=x_2$, es decir, sustituyendo las variables libres de una fórmula cuyo número de Gödel es x_1 por el número de Gödel x_1 . El resultado será una fórmula con número de Gödel x_3 . Ello se expresa así:

$$f_4(x_1, x_3) \quad \text{Ec. 28}$$

Lo siguiente que hay que hacer es construir esta proposición:

$$\neg \exists x_1 \exists x_2 (f_2(x_1, x_2) \wedge f_4(x_3, x_2)) \quad \text{Ec. 29}$$

Calculamos el número de Gödel de esta fórmula y sale NU.

Entonces si hacemos autosustitución de la única variable libre x_3 por NU queda la famosa proposición de Gödel:

$$\exists x_2 \neg \exists x_1 (f_2(x_1, x_2) \wedge f_4(\text{NU}, x_2)) \quad \text{Ec. 30}$$

Cuyo número de Gödel vamos a llamarlo G.

Y que significa que existe una fórmula con número de Gödel x_2 para la que no hay demostración. También dice que esa fórmula se obtiene al autosustituir NU sobre las variables libres de la ecuación 29. Dicho con otras palabras: la fórmula cuyo número de Gödel procede de la autosustitución de NU en la ecuación 29 no es demostrable.

Este formalismo se puede leer en español así:

Proposición G: “no es posible demostrar la proposición G” Ec. 31

Esta proposición quizás sea verdadera o quizás sea falsa. Examinemos cada caso:

- Si la proposición G es verdadera, entonces la proposición G no se puede demostrar, pero hemos supuesto que es verdadera. Por tanto, el sistema es incompleto.
- Si la proposición G es falsa, entonces sí se puede demostrar G. O sea, el sistema es incoherente.

Vaya lío ¿verdad? Hemos creado una proposición que es verdadera pero indemostrable; o es falsa y demostrable a la vez. Y esta contradicción la hemos logrado usando únicamente los números enteros, haciéndoles hablar de sí mismos. Las matemáticas no han sido las mismas desde entonces.

Ahora que ya hemos visto la demostración, quisiera hacer algunos comentarios adicionales.

Las contradicciones pueden parecer destructivas, pero son todo lo contrario. Son fuente de creación y de aprendizaje, y la ciencia progresa gracias a ellas. Por ejemplo, en la época en que solo existían los números naturales (positivos) y yo tenía 1000 en un banco y a continuación saqué 1500, aparentemente ello es una contradicción, pues es imposible de realizar físicamente. En la práctica, acabo de inventar los números negativos, es decir, acabo de pedir un préstamo. Lo mismo pasa con los números complejos y otras generalizaciones del concepto de número, como los tensores, los cuaterniones y los transfinitos. De la misma manera, cuando un experimento

de la física da lugar a una contradicción con las teorías existentes ello es un momento de júbilo, pues significa que se acaba de descubrir algo nuevo.

El teorema de Gödel es una versión sofisticada de la paradoja de Epiménides, el cretense, cuando dijo “Todos los cretenses son mentirosos”. La frase de Epiménides solo es verdadera si es falsa, y solo es falsa si es verdadera, lo cual nos lleva a una contradicción. Podemos ya intuir que para lograr contradicciones hay que tener la capacidad de autorreferencia, o sea, sistemas que hablen de sí mismos, como decíamos en el capítulo anterior. Te propongo el problema 8.

Problema 8: Dos autorreferencias divertidas

- Esta frase contiene _____ caracteres
- Esta frase no verbo

Antes del trabajo de Gödel ya se conocían otras paradojas lógicas. Seguramente la más famosa es la de Russell cuyo enunciado está en el problema 9, para que lo pienses un poco.

Problema 9: El barbero de Sevilla

En Sevilla existía un barbero que afeitaba a todas las personas que no se afeitaban a sí mismas. ¿Se afeitaba este barbero a sí mismo?

A veces tantas paradojas y tanto razonamiento contradictorio hacen que sea difícil entendernos a nosotros mismos, como en el problema 10 que está mal resuelto en varios libros.

El razonamiento llamado “diagonalización” que usó Gödel lo inventó Cantor, y es muy elegante y útil. Con él demostró que los números reales no son numerables, es decir, no se pueden poner en correspondencia biyectiva (uno a uno) con los números naturales. De hecho, la cardinalidad del conjunto de los números reales es un infinito mucho mayor que la de los números naturales. O sea, que hay una jerarquía de números infinitos.

El argumento de Cantor es más o menos así: sin perder la generalidad, vamos a listar todos los números reales que hay en el intervalo $[0,1)$ en cualquier orden astuto que uno desee o se haya inventado. Sin perder la generalidad, vamos a representarlos en binario y al primero le asignaremos el número natural 1, al segundo el 2, y así sucesivamente. Como están en binario tendrán la forma de $0.xxxxx$ donde las x son 0 o 1. Supongamos que ya tenemos la lista en correspondencia con los números naturales, como en el ejemplo de la figura 43.

Problema 10: Paradoja de Richard

Vamos a jugar con otra paradoja que habitualmente no se entiende bien. Para ello:

- Expresemos en español todas las propiedades de los números naturales de la aritmética (ej: "para todo número natural existe otro mayor que él", etc.).
- Ordenemos la lista de propiedades de forma lexicográfica (orden alfabético).
- Numeremos las propiedades (la primera es la número 1; la siguiente es la 2...).
- Como cada propiedad está asociada a un número entero, puede ocurrir que ese número cumpla con la propiedad.
- Ejemplo: si la propiedad 17 dice que "No ser divisible por ningún otro entero excepto por la unidad y por él mismo", entonces el número 17 si cumple con la propiedad 17.
- Ejemplo: si la propiedad 22 dice que "Ser producto de un entero por sí mismo", entonces el número 22 no cumple con la propiedad 22.
- Definición de nueva propiedad: Se llama Richardiano a todo número que no cumple con la propiedad que representa. Y no-Richardiano a los que si cumplen con su respectiva propiedad.
- Esta definición también estará en la lista, de modo que tendrá un número asociado R.

Número	Propiedad	¿Richardiano?
1		
2		
...		
R	No cumplir con la propiedad que representa	?
...		
17	No ser divisible por ningún otro entero excepto por la unidad y por él mismo	No
...		
22	Ser producto de un entero por sí mismo	Si
...		

¿Es R Richardiano?

```

0 . 0 0 0 1 1 0 1 0 0 ... → 1
0 . 0 0 1 0 0 1 1 1 1 ... → 2
0 . 0 0 1 1 0 1 0 1 0 ... → 3
0 . 0 0 1 1 1 0 0 0 0 ... → 4
0 . 0 1 0 0 0 1 0 1 0 ... → 5
0 . 1 1 0 1 0 1 1 1 0 ... → ...
...
    
```

Figura 43. Una posible numeración de los reales

Si ignoramos la primera cifra y el punto de los reales, podemos tomar las cifras de la diagonal y construir un nuevo número intercambiando ceros y unos. En la diagonal (Figura 44) está 001101... que cambiamos por

110010... Ponemos el “0.” delante, para que quede dentro del intervalo [0,1) y acabamos de construir un número binario que no está en la lista.

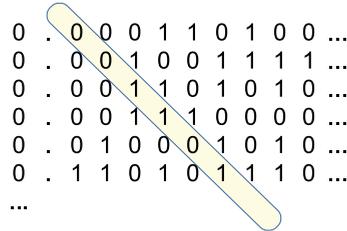


Figura 44. Tomamos la diagonal: 001101... en este ejemplo

Y no lo está porque no puede ser el primero de la lista, ya que la primera cifra decimal no coincide, por la construcción del número. Tampoco puede ser el segundo, porque la segunda cifra no coincide. Tampoco el tercero, pues no coincide la tercera cifra. Lo mismo pasa con los infinitos números de la lista; no puede ser ninguno de ellos, a pesar de que está en el intervalo [0,1). La conclusión es que la hipótesis de partida tiene que ser incorrecta: es imposible crear una lista con todos los números reales del intervalo [0,1).

Problema de la parada de Turing

Habitualmente se diseñan los sistemas axiomáticos de modo que no tengan contradicciones internas pues ello es lo peor que puede ocurrir: si aplicando una secuencia de reglas una proposición sale verdadera, mientras que aplicando otra secuencia la misma proposición sale falsa, entonces cualquier cosa se puede demostrar en ese sistema. En el recuadro 3 se muestra por qué⁴⁰.

Recuadro 3: En los sistemas axiomáticos incoherentes se puede demostrar cualquier cosa

- Supongamos que tenemos una contradicción en mi sistema, es decir, $(A \wedge \neg A)$ es verdadero.
- De allí se deduce que A es verdadero y que A es falso, simultáneamente.
- Si A es verdadero, también lo es $(A \vee B)$ siendo B cualquier proposición.
- Como $(A \vee B)$ es verdadero y A es falso, entonces B es verdadero.

Acabamos de demostrar que cualquier proposición B es verdadera. Q.E.D.

40 De todos modos, hay que recordar que existen las lógicas paraconsistentes, donde se permiten cierto tipo de contradicciones (por ejemplo, en las lógicas difusas o las multivaluadas o las temporales) y donde el sistema no se derrumba sino que tiene mucha utilidad.

Entonces, para evitar la incoherencia se opta por aceptar la incompletitud como un mal menor. Por eso, Turing se preguntó si era posible emplear un algoritmo para detectar en un sistema axiomático cuáles proposiciones son indecidibles, para filtrarlas fuera y quedarse con las demás. Empleando su definición de Máquina de Turing Universal demostró que no era posible hacerlo.

En vez de una demostración formal, vamos a verlo con un programa en Ruby, que es más sencillo (Recuadro 4).

Esto tiene consecuencias importantes a la hora de desarrollar *software*: dado que es imposible saber de antemano en un caso general y de forma automática si un programa va a llegar a su final o va a quedarse atrapado en un bucle infinito, entonces tampoco es posible saber si contiene errores. Esto se expresa mejor con una variante de la Ley de Murphy: “Todo programa contiene errores hasta que no se demuestre lo contrario, lo cual es imposible”.

Recuadro 4: Demostración del Halting Problem

Hipótesis: se puede escribir una función parada(), que lea un programa y su entrada, lo analice y retorne true si detecta que el programa se va a detener, y false en caso contrario.

```
def parada(programa, entrada)
  # Aquí viene un programa maravilloso,
  # pero los detalles no nos interesan
end
```

Suponiendo que eso es posible, entonces yo puedo escribir el siguiente programa, al que llamaré diagonal.rb:

```
programa = gets #Lee la entrada y la almacena en un string
if parada(programa, programa)
  loop do      #bucle infinito
  end
else
  return true
end
```

Por último, desde el shell de Linux ejecuto:

```
./diagonal.rb < diagonal.rb
```

Y entonces puede ocurrir una de estas dos cosas:

- Si la función parada(programa, programa) retorna true (recordemos que es una función que tú me diste y que me juraste que era capaz de detectar programas que paran, retornando true, y que no paran, retornando false), entonces mi programa diagonal.rb se queda en un bucle infinito, es decir, no para. Por tanto, tu función no funciona bien, pues se equivoca.
- Si la función parada(programa, programa) retorna false, entonces mi programa diagonal.rb retornará true y finalizará su ejecución. De nuevo, tu función se equivocó al predecir que no iba a terminar.

Por tanto, la hipótesis inicial de que era posible construir la función parada() es falsa. **Q.E.D.**

Por ello, las metodologías ágiles de programación guiadas por pruebas (como TDD) son muy buenas. Primero se escriben las pruebas, luego se escribe el programa hasta que pase todas las pruebas. Y cada vez que un usuario reporte haber encontrado un nuevo error, se escriben las pruebas correspondientes que, obviamente, el programa no puede pasar. Después se corrige el programa hasta que pase estas nuevas pruebas y también todas las pruebas anteriormente escritas. Con ello se garantiza que antiguos errores ya corregidos no vuelvan a la vida como consecuencia de los cambios realizados para corregir el nuevo error. Podemos así decir que el número de errores del programa irá decreciendo asintóticamente a cero, conforme pasa el tiempo.

Otra alternativa para garantizar programas libres de errores es evitar computación completa, básicamente impidiendo por diseño que aparezcan bucles infinitos (cualquier tipo de bucle infinito, iterativo o recursivo). Esto quita mucha expresividad a cualquier lenguaje de programación, pero de este modo permite en teoría analizar un programa antes de ejecutarlo, en busca de errores. Eso es lo que se plantea, por ejemplo, en las metodologías que usan precondiciones, postcondiciones e invariantes para cada función. O en las metodologías que parten de una especificación matemática para traducirla a un programa.

De alguna manera ello divide la computación en dos (Figura 45): si quieres un lenguaje de programación (o una metodología) con garantías de producir programas libres de errores entonces no puedes aspirar a tener computación completa. Y si quieres inteligencia artificial, que de algún modo implica creatividad, entonces no puedes limitarte a tener una computación no-completa. Se necesita computación completa o incluso más.

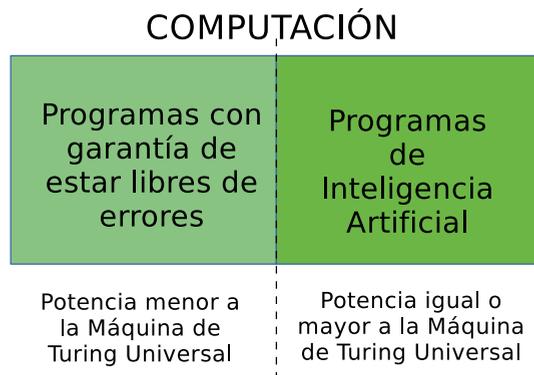


Figura 45. Las dos facetas de la computación

Entonces esto es un límite computacional: para lograr programas de inteligencia artificial hay que renunciar a que estén libres de errores. También se puede ver al revés, de una forma más optimista: los errores son una forma de creatividad. O, dicho más claramente, no se puede esperar verdadera inteligencia sin el riesgo de cometer errores, de equivocarse, de llegar a contradicciones.

Cuando pensamos que el ser humano es inteligente no nos damos cuenta de ello: somos falibles, nos dejamos engañar con facilidad, nos equivocamos con las sumas, lo que decimos un día lo desdecimos al día siguiente, razonamos mal con la lógica. Lo que realmente queremos decir es que la humanidad es inteligente, es decir, el proceso de tener muchos agentes que, colectivamente, van depurando sus errores, eliminando sus contradicciones, seleccionando las verdades (temporales) que mejor resultado les da, y transmitiéndolas por escrito a futuras generaciones. A ese proceso se le llama ciencia.

Complejidad computacional

La complejidad computacional, tanto temporal como espacial, es un tema bien conocido para cualquier científico de la computación. Hagamos un repaso rápido: se definen los **problemas de decisión** como aquellos que tienen una respuesta booleana (sí o no), mientras que los **problemas de optimización** buscan un óptimo entre un conjunto posiblemente infinito de candidatos. Un problema de optimización puede ser el del vendedor viajero: dado un conjunto de ciudades, averiguar en qué orden hay que visitarlas todas para minimizar la distancia recorrida. Y el problema equivalente de decisión es preguntar si existe un orden de recorrido tal que su distancia sea menor de 56 km.

Para resolver los problemas se usan **modelos de cómputo**. El más usado hasta ahora es la Máquina de Turing Universal. Pero existen otros modelos tan fantásticos como la hipercomputación, que se caracteriza porque sus computadores pueden trabajar con números reales en tiempo finito⁴¹. O tan a punto de convertirse en realidad como la computación cuántica⁴², donde, aprove-

41 Por ejemplo, un computador cuya frecuencia de reloj se duplique cada cierto tiempo. Podríamos calcular así la primera cifra de π en un tiempo T , la segunda en $T/2$, la tercera en $T/4$, etc. Dado que la serie $T+T/2+T/4+\dots$ converge a $2T$, habremos calculado un número real trascendente en un tiempo finito. Obviamente ese computador nunca se podrá construir en nuestro universo, porque al aumentar la frecuencia del reloj tropezaríamos primero con limitaciones tecnológicas y luego físicas.

42 Ya hay varios computadores cuánticos trabajando muy bien en varios laboratorios del mundo, y falta poco para que se vendan al público, pero lo que definitivamente escasea son los algoritmos que puedan correr de forma efectiva en estas máquinas.

chando el fenómeno de superposición, se puede trabajar simultáneamente con un número enorme de candidatos a solución. Todavía no está claro si la computación cuántica es equivalente a la Máquina de Turing Universal, aunque más rápida, o si se podrá lograr hipercomputación, o si dará lugar a una clasificación intermedia de potencia de cómputo. Hay muchos artículos hablando de ello, pero son contradictorios. Lo que sí está claro es que dado que la inteligencia artificial depende fuertemente del modelo de cómputo, con la llegada de la computación cuántica todo va a cambiar drásticamente y es inminente que los computadores superen a los humanos en cualquier tarea.

A su vez cada modelo de cómputo da lugar a una clasificación de problemas: problemas tratables y problemas intratables (imposibles de resolver en un tiempo razonable). La clasificación más estudiada es la que se hace para las Máquinas de Turing Universales, cuyas clases de problemas más conocidas son (siendo N el número de bits necesarios para codificar los datos de entrada del problema en cuestión usando un código razonable):

- Problemas P (polinomiales): aquellos que se resuelven aplicando un algoritmo determinista en un tiempo máximo proporcional a un polinomio de N ($T \propto N^K$).
- Problemas NP (polinomiales no-deterministas): aquellos que se resuelven aplicando un algoritmo no-determinista en un tiempo máximo proporcional a un polinomio de N . También se pueden definir como problemas EXP pero verificables en tiempo polinomial, es decir, si tengo un candidato a solución, para verificar si realmente lo es me demoro un tiempo polinomial.
- Problemas EXP (exponenciales): aquellos que se resuelven aplicando un algoritmo determinista en un tiempo máximo proporcional a una exponencial de N ($T \propto K^N$).

Ejemplos de problemas P son buscar un elemento concreto en una lista desordenada de N elementos (el tiempo de búsqueda es proporcional a N). Ordenar los elementos de una lista también es un problema polinomial.

Un ejemplo de problema NP es el de decisión del vendedor viajero. La idea es que si tenemos N ciudades hay $N!$ formas distintas de recorrerlas, lo cual crece más rápido que una exponencial. Sin embargo, si me dan una secuencia ordenada de ciudades, me demoro muy poco (solo tengo que hacer N sumas, es decir, la verificación es polinomial) en saber si el recorrido es menor que una cierta distancia dada. Otro problema NP es el de satisfabilidad: dada una función lógica de N variables (de álgebra de Boole) averiguar que valores (0 o 1) deben tomar las variables para que la fórmula ofrezca como resultado 1. Por ejemplo:

$$f = (x_1 + x_2 + x_3) \cdot (x_1 + x_3) \cdot x_3 \quad \text{Ec. 32}$$

$$g = (x_1 + x_2) \cdot (x_1 + x_3) \cdot x_1 \cdot (x_2 + x_3) \quad \text{Ec. 33}$$

La función f se satisface con $x_1=0$, $x_2=1$, $x_3=1$ mientras que la función g no es satisfactible. Pero para averiguarlo hay que explorar todas las combinaciones posibles, que son 2^N . Pero dada la solución es muy rápido verificarla, en tiempo polinomial.

Un ejemplo de problema EXP es averiguar cuál es la mejor jugada que puedo hacer ante un tablero dado de ajedrez. La única forma que tengo de averiguarlo es probando todas las jugadas posibles hasta llegar al final de la partida. Y aunque me digan cuál jugada es la correcta, para verificarlo también debo de probar todas las jugadas posibles a partir de ahí, de modo que la verificación no es polinomial.

Hay muchas más clases de problemas pero solo nos fijaremos en estas tres pues son las que más frecuentemente aparecen en problemas prácticos reales actuales.

Se han encontrado muchísimos problemas NP en campos muy diversos: teoría de conjuntos, teoría de grafos, búsqueda en árboles, conectividad, trazado de rutas, particiones, compresión de datos, bases de datos, planificación de tareas por sistemas operativos, teoría de números, juegos, lógica proposicional, lenguajes formales, autómatas, etcétera. Los más usados en ámbito académico, aparte del de satisfabilidad y del viajero son el empaquetamiento de la mochila⁴³, y la asignación de aulas y horarios en una universidad.

Algo muy interesante: se dice que un problema es NP-completo si existe un algoritmo de tiempo polinomial que lo transforma al problema de la satisfabilidad. Los problemas NP-completos son los más complejos dentro de la clase NP. Y esto implica que todos los problemas NP-completos son equivalentes en el sentido de que si se logra solucionar uno de ellos en tiempo polinomial, entonces todos los demás también quedan resueltos. Y también lo contrario: si se demuestra que un problema NP-completo es imposible solucionarlo en tiempo polinomial, entonces es imposible para todos ellos. Hay que advertir que no es lo mismo resolver una instancia de un problema que resolver el problema general, que es lo que nos interesa aquí. Por ejemplo, hay problemas de satisfabilidad muy sencillos de resolver como $f=x_1+x_2$, pero ello no nos vale. Lo que estamos buscando es la solución al problema general. Entonces se sabe que $P \subseteq NP$, pero no se sabe si ambos conjuntos

43 Dada una mochila con una cierta capacidad L y un conjunto de N objetos de tamaños $\{c_1, c_2, \dots, c_N\}$ y con unos valores $\{v_1, v_2, \dots, v_N\}$ se trata de empaquetar los objetos que se puedan sin sobrepasar la capacidad de la mochila y maximizando la suma de sus valores.

coinciden, es decir, si existe un truco, un algoritmo, para resolver en tiempo polinomial los problemas NP. Ello es uno de los llamados “problemas del milenio”, que cuentan con un premio de un millón de dólares a quien lo logre resolver.

Quien no conozca los detalles sobre las clases de complejidad computacional puede remitirse a Garey y Johnson (1979). De todos modos, lo que quiero resaltar ahora es la inutilidad de enfrentar problemas NP y EXP aumentando la potencia de cómputo. Obviamente, incluso los problemas EXP son resolubles para valores de N muy pequeños. Pero en cuanto N crece se vuelven intratables, incluso si la tecnología de cómputo avanza sustancialmente. Para entenderlo, en la figura 46 comparamos varios problemas de complejidad temporal N , N^2 , N^3 , 2^N y 3^N , y suponemos que ya hemos logrado solucionarlos para una entrada de tamaño N_1 . Entonces, si esperamos unos años hasta que aparezcan computadores 100 veces más rápidos y luego 1000 veces más rápidos, para los problemas polinomiales vemos que merece la pena, mientras que para los intratables (NP y EXP) es muy poco lo que se lograr mejorar. Eso es lo que significa que un problema sea intratable. Y ello es un límite computacional.

CLASE	COMPLEJIDAD	SOLUCIÓN ACTUAL	SOLUCIÓN CON UN COMPUTADOR 100 VECES MÁS RÁPIDO	SOLUCIÓN CON UN COMPUTADOR 1000 VECES MÁS RÁPIDO
P	N	N_1	$100 N_1$	$1000 N_1$
P	N^2	N_1	$10 N_1$	$31,6 N_1$
P	N^3	N_1	$4,64 N_1$	$10 N_1$
NP, EXP	2^N	N_1	$N_1+6,64$	$N_1+9,97$
NP, EXP	3^N	N_1	$N_1+4,19$	$N_1+6,29$

Figura 46. Diversos problemas y lo que se gana al aumentar la potencia de cómputo

Mientras tanto, hay alternativas para tratar de resolver los problemas NP:

- Modificarlo simplificándolo para convertirlo en P.
- Buscar un algoritmo que corre en tiempo polinomial la mayoría de las veces, aunque no tenga garantías de hacerlo siempre.
- Conformarse con soluciones no-óptimas. Para ello se han desarrollado herramientas que realizan búsquedas heurísticas, como un “genio mágico”, un oráculo que predice cual puede ser la solución, que debe luego verificarse aprovechando esta característica de los problemas NP. Algunos de estos algoritmos son las redes neuronales y los algoritmos evolutivos. De alguna forma los algoritmos evolutivos (que vere-

mos en el capítulo con el mismo nombre) están bien adaptados a los problemas NP pues generan candidatos a solución por medio de heurísticas generales (operadores de reproducción) y verifican su bondad con una función de aptitud que es rápida de evaluar.

Teorema del No-Free-Lunch

El teorema de *No-Free-Lunch* (NFL) fue ideado en 1992 por Wolpert y Macready, como un marco donde investigar los problemas de búsqueda, enfocándose en la conexión entre funciones de aptitud y efectividad en la búsqueda. Lo que dice este teorema es: aplicando algoritmos de búsqueda a todos los tipos de problemas existentes, todos los algoritmos se comportan igual en promedio.

Concretamente, si tomamos un algoritmo de búsqueda, se comportará bien (mejor que el promedio) en la mitad de los problemas y mal (peor que el promedio) en la otra mitad. Es importante considerar el universo de todos los problemas posibles, pues de otro modo el NFL no tiene nada que decir.

También se puede formular así (Figura 47): por cada par de algoritmos de búsqueda hay tantos problemas en el que el primer algoritmo es mejor que el segundo como problemas en el que el segundo algoritmo es mejor que el primero. O sea que si comparamos algoritmos genéticos con enfriamiento simulado o incluso con búsqueda aleatoria, y el algoritmo genético se comporta mejor en cierto tipo de problemas, siempre habrá otro tipo de problemas donde se comporte peor.

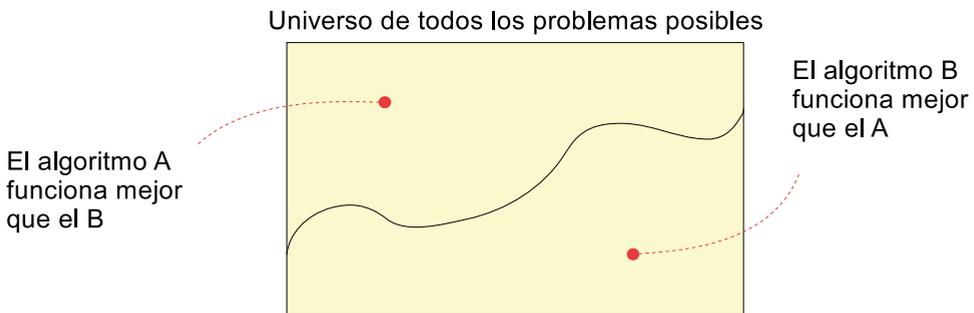


Figura 47. Explicación gráfica del NFL

Una consecuencia es que si no introducimos ningún conocimiento suplementario del dominio del problema en nuestro algoritmo, es tan probable que trabaje peor que la búsqueda estocástica, como que trabaje mejor. Esto es cierto para cualquier tipo de algoritmo de búsqueda. De modo que no se debe elegir un algoritmo basándose únicamente en lo bien que funcionan en

artículos de investigación, sino que es necesario conocer su comportamiento en el dominio del problema que vayamos a tratar (Figura 48).

Y esto debe ser un llamado de advertencia para quien diseña nuevos algoritmos: no basta con mostrar que tu algoritmo es mejor que todos los demás en un caso particular, ya que siempre es posible encontrar ese caso para cualquier algoritmo, aun cuando sea un algoritmo pésimo. Lo importante es mostrar que funciona en un amplio número de problemas, ojalá con relevancia en el mundo real. Por ejemplo, si estás diseñando un nuevo algoritmo para el vendedor viajero, no basta compararlo con otros para un único conjunto de ciudades, ni siquiera para 10 conjuntos. Es más sensato hacerlo con millones de conjuntos de ciudades cuyas posiciones se generen al azar.

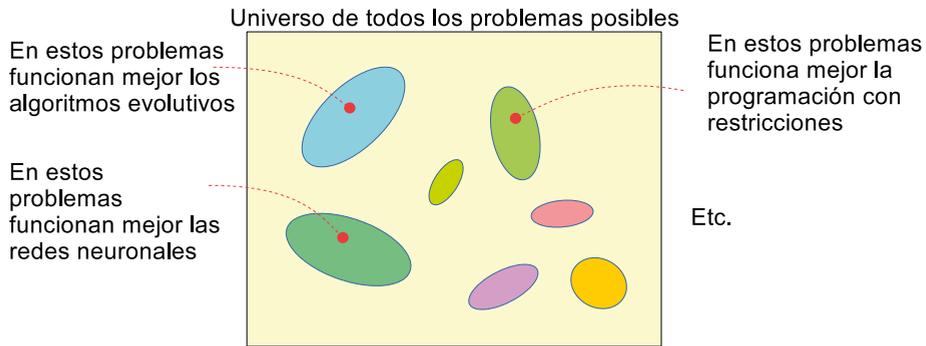


Figura 48. Dominios de problemas donde funciona mejor cada algoritmo

Volviendo al tema, resulta que el teorema NFL aparentemente echa por tierra la esperanza de conseguir el gran algoritmo inteligente que sirva para resolver todo tipo de problemas. Esta limitación computacional parece impedir que se pueda algún día alcanzar la verdadera inteligencia artificial. Porque por más que nos esforcemos en diseñar algoritmos inteligentes, fallarán para la mitad de los problemas que les planteemos. Ello es decepcionante.

En la ciencia ocurre algo similar: las teorías más generales son las que explican menos cosas, y las más particulares las que pueden dar predicciones más detalladas⁴⁴.

Sin embargo, algo hay mal interpretado en el NFL porque, por un lado, existen muchos algoritmos (evolutivos, *deep learning*) que funcionan muy bien para resolver problemas en muchas áreas. Además, el NFL conduce a una paradoja: si todo algoritmo requiere información adicional introducida por un humano para afinarlo en un dominio concreto donde funcione bien,

⁴⁴ Como chiste también se dice que los ingenieros saben de todo un poco, mientras que los PhD saben muchísimo de nada.

eso sitúa a la inteligencia humana de una manera implícita por fuera de los algoritmos. Está diciendo que la inteligencia humana no se puede modelar por medio de ningún algoritmo, ni los clásicos deterministas, ni aún por los no-deterministas. Y si esto es así, se acaba no solo con la inteligencia artificial en sentido “fuerte” sino también con la “débil”. Y le daría un sentido “mágico” a la inteligencia humana, que no estamos dispuestos a aceptar aquí (Figura 49).



Figura 49. Este nuevo algoritmo (*Humano + Algoritmo X*) ¿está libre del NFL?

Para salir de esta paradoja tenemos que pensar si hay algo mal interpretado en los NFL. ¿Quizás los problemas que habitualmente tratamos de resolver están sesgados y no abarcan toda la gama de problemas posibles? ¿Es el mundo en que vivimos el que está sesgado? ¿Es nuestra percepción la sesgada? Posiblemente, la respuesta sea “sí” a todas estas preguntas, ya que el mundo es un producto de la evolución natural, por lo que dentro de él no se dan todos los problemas posibles.

En este sentido, el filósofo Malcolm R. Forster (2002) plantea el siguiente experimento mental: supongamos que hay un universo distinto al nuestro, que dura exactamente 2 días y en cada día existe exactamente 1 objeto, que puede ser una esfera o un cubo. El objeto podría ser el mismo o distinto, en cada uno de los 2 días. Entonces, ese universo tiene 4 posibles historias (Figura 50):

	Historia 1	Historia 2	Historia 3	Historia 4
Primer día	esfera	esfera	cubo	cubo
Segundo día	esfera	cubo	esfera	cubo

Figura 50. Posibles historias del universo de Forster

Hay exactamente 4 posibles algoritmos que intentan predecir lo que ocurrirá en el segundo día, sabiendo lo que ocurrió en el primer día:

- Igual: en el segundo día habrá el mismo objeto que en el primer día.
- Diferente: habrá un objeto distinto.
- Esfera: habrá una esfera, independientemente del primer día.
- Cubo: habrá un cubo, independientemente del primer día.

La probabilidad de las cuatro historias la suponemos igual (1/4). Por tanto, la probabilidad de acierto de cada algoritmo predictor es:

- Igual: 1/2
- Diferente: 1/2
- Esfera: 1/2
- Cubo: 1/2

Tal como indica el NFL, no hay ningún algoritmo privilegiado. En el conjunto global de todos los problemas posibles, todos los algoritmos se comportan en promedio igual.

Sin embargo, en nuestro universo real, sabemos que hay “uniformidad en la naturaleza”, que se traduce a que las cosas no aparecen y desaparecen por si solas, sin ninguna causa. Ello implica que las únicas historias posibles son la 1 y la 4. Nuestro universo está “sesgado”. Y entonces, la probabilidad de acierto de cada algoritmo es:

- Igual: 1
- Diferente: 0
- Esfera: 1/2
- Cubo: 1/2

En este universo sesgado, el algoritmo “igual” es el mejor de todos. Y si el algoritmo de predicción hubiera sido el resultado de un proceso de búsqueda evolutiva, no habría sido necesario proporcionarle ninguna información adicional para que encontrase la solución a cómo funciona nuestro universo.

Es decir, el NFL es inviolable, pero probablemente nunca tengamos que enfrentarnos a resolver todos los problemas posibles. Gracias a ello es que funciona nuestra inteligencia humana, y nada impide que lleguemos a lo mismo con inteligencia artificial. En ambos casos siempre habrá problemas que no se puedan resolver (por ejemplo, descifrar algoritmos de criptografía fuerte).

¿QUÉ IMPORTANCIA TIENEN ESTOS LÍMITES?

Durante mucho tiempo los principales filósofos pensaban que la mente humana estaba exenta de estas limitaciones debido a que había algún elemento misterioso no sometido a las leyes físicas (Searle, 1997) o a las computacionales (Penrose, 1991, 1994). La ciencia avanza y estas ideas han sido descartadas. Lo más bonito es que ahora podemos investigar y entender cómo el cerebro lidia con estas limitaciones. Vamos a discutir varios ejemplos.

En cualquier empresa hay que asignar recursos a personas para que realicen tareas. Por ejemplo, en las universidades hay que asignar aulas a profesores en determinados horarios, para que puedan dar allí sus asignaturas. Incluso sin tener en cuenta que también hay que asignar estudiantes a cada asignatura, el problema es NP y tiene varias restricciones. Las principales son: cada asignatura tiene un conjunto limitado de profesores que la pueden dar; un profesor no puede estar simultáneamente en dos aulas; y un aula no puede ser asignada a varias asignaturas en horarios que se solapen. Para resolverlo, básicamente hay que probar todas las combinaciones posibles de aulas, asignaturas, profesores y horarios, e ir descartando las que no cumplen con alguna de las restricciones. El objetivo principal a satisfacer son las asignaturas. Conforme el número de asignaturas crece, el tiempo para calcular una solución aumenta exponencialmente. Y no hay garantías de que existan soluciones. Pero, además, ¿de qué sirve saber que hay una solución si el computador va a demorarse 5 años en encontrarla? Y si la tarea se le encomienda a un humano, va a ser mucho más.

La verdad es que sí hay una forma de solucionar problemas cuyo tiempo de cálculo es exponencial, gracias a que las exponenciales al principio tienen un comportamiento muy suave, casi lineal, de aumento casi imperceptible, como se puede ver en la figura 51. Es la función e^n , y allí nos da la impresión de que hasta $n=96$ pareciera que no está pasando nada.

Por ejemplo, la globalización es un fenómeno exponencial que lleva ocurriendo prácticamente desde que se formó la Tierra hace 4.500 millones de años, siendo los hitos humanos recientes más significativos la salida del *Homo Sapiens* de África hacia Europa y Asia, así como el salto hacia América. Y luego en sentido inverso: la colonización de África, la comunicación con Asia, el descubrimiento de América y la invención de Internet. Pero solo después de que ocurriera este último acontecimiento es que hemos sido conscientes de lo que significa el fenómeno que permite comunicaciones instantáneas con cualquier persona del globo, interacciones económicas y políticas desde y hacia cualquier rincón del planeta, tener noticias de lo que ocurre en cualquier parte así como poder influir también en ello⁴⁵, que hasta entonces pasó desapercibido y ni siquiera tenía nombre. Eso es una exponencial. Otro ejemplo más dramático es el calentamiento global, que también es exponencial, pero que nadie lo advierte porque estamos en una fase de subida del mar de unos milímetros cada año, lo cual es imperceptible. La ciencia nos lo cuenta porque tiene equipos sofisticados para medirlo, pero

45 Usando plataformas como Avaaz.org o Change.org, que se están configurando como los nuevos actores de la política global.

no parece que la gente le dé la importancia debida. Y para cuando seamos conscientes de la gravedad del asunto, será demasiado tarde.

Volviendo al problema de asignación de aulas, horarios y profesores a asignaturas, se puede solucionar fácil y rápidamente si tenemos mucha holgura, de modo que los tiempos de cómputo sean rápidos porque hay muchísimas soluciones posibles. Holgura en este caso significa tener muchas aulas y profesores, dado que los horarios no se pueden estirar (en una semana se puede disponer de 6 días laborables x 14 horas hábiles/día = 84 horas), siendo imposible físicamente superar las 168 horas/semana. De modo que la única solución es contratar más profesores y construir más aulas. Pero entendamos bien qué significa este concepto de holgura.

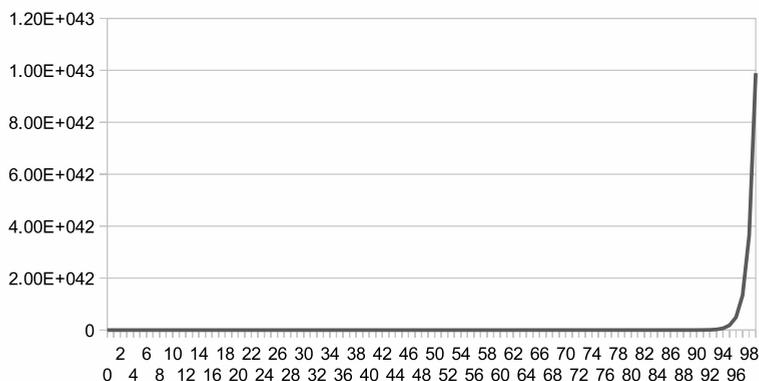


Figura 51. Cien primeros valores de e^n

Si no hay holgura, el problema no tiene solución. Por ejemplo, si hay muchos profesores pero solo hay 2 aulas, y el número de asignaturas es 140 (suponiendo que cada asignatura dure 3 horas/semana), entonces se requieren $3 \cdot 140 / 2$ horas/semana = 210 horas/semana. Como ya dijimos, la semana no tiene tantas horas. Es imposible asignar todas las asignaturas.

Si construimos muchas aulas, por ejemplo 10, habrá mucha holgura: Los cálculos ahora nos dan $3 \cdot 140 / 10$ horas/semana = 42 horas/semana. Dado que la semana dispone de 84 horas, tenemos un 50% de aulas desocupadas a lo largo de la semana. Hay una holgura del 50%. Y eso, que puede parecer ineficiente, es muy bueno desde el punto de vista de la efectividad de asignar las aulas.

También podemos verlo de esta manera: si tenemos un tiempo fijo limitado para solucionar un problema, conforme las variables restrictivas

del problema aumentan (número de estudiantes, número de asignaturas), la cantidad de posibilidades que se deben analizar crece exponencialmente (Figura 52).



Figura 52. Crecimiento exponencial de las combinaciones a analizar

Los científicos de la computación saben que si el tiempo de cómputo es limitado, como siempre ocurre, el número de combinaciones que se alcancen a explorar será una fracción muy pequeña del total. De modo que la probabilidad de encontrar una solución decrece bruscamente a cero al superarse un cierto umbral en el tamaño del problema (Figura 53).



Figura 53. Probabilidad de encontrar una solución con tiempo de cómputo limitado

Un gerente observa lo mismo pero desde otro punto de vista: la probabilidad de encontrar una solución aumentará bruscamente cuando la holgura supere un cierto umbral (Figura 54), que es el mismo de la exponencial. En problemas NP esta transición es muy brusca. Tener o no tener holgura (por encima o por debajo de un cierto umbral) produce un cambio de fase entre lograr o no lograr solucionar el problema. Este cambio de fase es típico de sistemas complejos, de modo que los problemas NP generan una complejidad sistémica y no solo algorítmica.

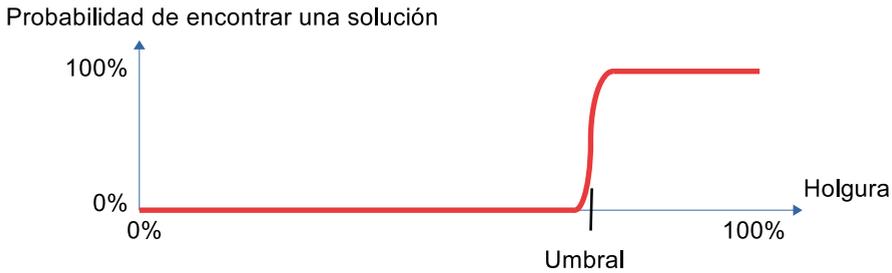


Figura 54. Probabilidad de encontrar solución, en función de la holgura que haya

Podemos calcular la holgura en este caso como resta de dos rectángulos (Figura 55).

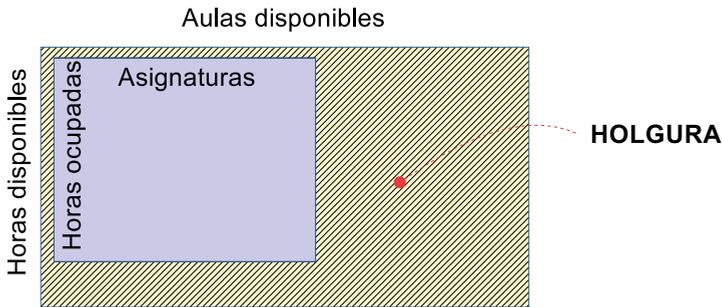


Figura 55. $Holgura = (\text{Rectángulo disponible} - \text{Rectángulo ocupado}) / \text{Rectángulo ocupado}$

Pero eso no es suficiente para entender la holgura. En problemas combinatorios como este, un 50% de holgura (como la que hay en la figura 56) puede ser poco para solucionarlo y ello depende de la complejidad de los detalles. Se requerirá más holgura si hay muchas restricciones, si las aulas no son todas del mismo tamaño, sino que hay mucha diversidad, si las asignaturas no ocupan todas la misma cantidad de horas, si hay asignaturas compartidas por más de un profesor, si algunas sesiones no se dan en el aula sino en otro espacio (laboratorio), etc.

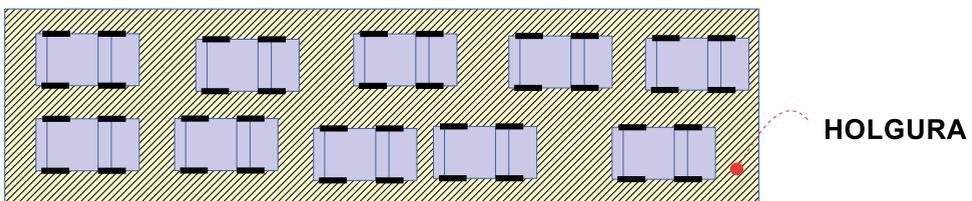


Figura 56. Una calle con automóviles. Hay congestión a pesar de que la holgura parece alta

Para entender por qué un modelo de eficiencia basado en rectángulos no es suficiente, consideremos un problema más intuitivo: el tráfico en una ciudad. Un alcalde que no quiera gastar en construir más calles podría preguntar el área asfaltada de la ciudad, el número de automóviles que circula por allí y el área que ocupa cada automóvil. Supongamos que son respectivamente 2 000 000 m², 100 000 automóviles y 10 m². Este alcalde podría decir que en su ciudad caben $2\,000\,000\text{ m}^2 / 10\text{ m}^2 = 200\,000$ automóviles. Y que nadie debe quejarse, pues en este caso hay solo la mitad, o sea, que caben todavía 100 000 más. ¡Pobre ciudad si tiene un alcalde así! No entiende la importancia de la holgura y, obviamente, nadie desearía vivir en esa ciudad tan congestionada.

Incluso en un estacionamiento público donde la mayoría de los vehículos están detenidos y no necesitan holgura delante para poder frenar, se requiere que alrededor de un 50% del área disponible esté libre para que los automóviles y peatones puedan movilizarse y abrir las puertas. Otro ejemplo similar lo vemos en la figura 57, que muestra un atasco habitual en nuestras ciudades donde, a simple vista, podemos estimar que hay todavía una holgura de más o menos un 50% entre los automóviles. Y, a pesar de ello, hay congestión. Si disminuyéramos la holgura al 0% nadie podría abrir las puertas de su vehículo ni nadie podría avanzar. Los automóviles quedarían en una especie de fase sólida para siempre.



Figura 57. Congestión vehicular en la Puerta de Alcalá de Madrid

Por tanto, que haya holgura es importante para que los sistemas funcionen y sean computables. Y nunca debería verse la holgura como un desperdicio de recursos, sino más bien como la flexibilidad que requiere un sistema para operar normalmente e incluso para poder reaccionar rápidamente ante imprevistos.

¿Cuánta holgura se requiere? Depende de cada sistema, y suele ser difícil de calcular. Es más rápido hallarlo experimentalmente. Sin embargo, en algunos casos sí se conoce teóricamente cuál es el umbral donde una holgura menor hace el sistema inviable. Por ejemplo, otro recurso muy importante que se suele asignar en cualquier empresa es el tiempo. Típicamente cada empleado debe hacer un conjunto de tareas, la mayoría de las cuales son periódicas y se conocen de antemano. Así, un profesor universitario debe dar clases, atender reuniones administrativas, participar en grupos de investigación, asistir a conferencias y otras responsabilidades. Muchas de estas tareas tienen periodicidad semanal, semestral o anual. Los algoritmos de planeación de tareas se conocen desde los albores de la computación (pues el sistema operativo sufre del mismo problema de asignar tiempos de *CPU* a tareas), y el primer límite teórico de lo que se puede lograr fue calculado por Liu y Layland en 1973: si todas las tareas son periódicas y conocidas de antemano, y si no hay restricciones entre ellas (recursos compartidos, prioridades, etc.) entonces lo máximo que se puede ocupar una *CPU* es un 69%. Es decir, es imposible dar garantías si la holgura es menor al 31%. Remarquémoslo: debemos conformarnos con que haya un 31% del tiempo disponible que no se use para nada (o como mucho para tareas de baja prioridad que no importe si se quedan sin hacer), si queremos que las tareas asignadas se ejecuten sin problemas. Y este es el mejor caso. Conforme aumentan las restricciones, la holgura que debe dejarse puede ser mucho mayor. Lo que Liu y Layland obtuvieron para *CPUs* de computador es igualmente aplicable a personas, de modo que tratar de planificar el 100% del tiempo de una persona en tareas de obligatorio cumplimiento es un disparate. Lo malo es que los administradores de empresas no suelen conocer estos resultados de ciencias de la computación.

La habitación china

El filósofo y lingüista John Searle y el matemático y filósofo Roger Penrose son los máximos exponentes de una corriente que usa el teorema de Gödel como prueba de que las máquinas jamás podrán ser inteligentes.

Como ambos atacan el funcionalismo y cualquier teoría computacional de la mente, vamos a analizar aquí sus ideas y a desmontarlas.

Su argumento es que el teorema de Gödel limita lo que un formalismo (que ellos identifican correctamente con un programa de computador determinista) puede llegar a hacer. Mientras que, para ellos, la mente humana no sufre de esas limitaciones. Por tanto, los programas de computador nunca podrán llegar a ser tan inteligentes como los humanos. Lo más que pueden hacer es simular algunos procesos mentales, lo que se llama “inteligencia

artificial débil” como jugar muy bien ajedrez. Pero tener pensamientos y sensaciones humanas (lo que se llama “inteligencia artificial fuerte”) está por fuera de sus posibilidades.

A Searle y a Penrose se les puede criticar tres cosas:

- Identifican los formalismos de Gödel con programas de computador, lo cual es correcto siempre y cuando esos programas no incluyan aleatoriedad. Es decir, es válido para formalismos computacionales como los programas en LISP. Pero cuando se incluye aleatoriedad, por ejemplo, permitiendo que una proposición sea a veces verdadera y a veces falsa, o que coexistan proposiciones afirmando algo y su contrario simultáneamente (como ocurre con los cromosomas de un algoritmo evolutivo) el teorema de Gödel no limita las verdades alcanzables. Dicho de otra manera, un formalismo que acepte contradicciones internas de una manera creativa y no destructiva, evita el límite de las verdades inalcanzables.
- Ambos pensadores se sienten muy inteligentes. Se miran en su interior y no creen que tengan ninguna limitación. Eso es un vano orgullo. Todas las personas tenemos limitaciones y los grandes desarrollos en matemáticas, ciencias y tecnología se los debemos al conjunto de la humanidad, a los vivos y a los ya fallecidos. No hay una única persona que pueda presumir de entender y potencialmente haber podido llevar a cabo cualquier avance científico. Y la ciencia funciona de manera similar a los algoritmos evolutivos, si hacemos el símil de una persona igual a un cromosoma. Entre los científicos hay contradicciones y hay cambios de opinión constantemente. No se puede decir que sean modelables con un formalismo matemático en el cual, una vez establecida la verdad, es verdad para siempre, como cuando decimos que $2+2=4$. Lejos de eso, una vez establecida una verdad científica (llamada teoría), no suele pasar mucho tiempo hasta que alguien encuentra un experimento que no casa, y hay que sustituirla por otra teoría más sofisticada. El teorema de Gödel es completamente aplicable a cada humano y también a la humanidad, y para alcanzar nuevas verdades tenemos que aceptar la falibilidad y las contradicciones.
- Como repetiremos varias veces a lo largo del libro, no hay forma de distinguir lo real de lo simulado, siempre que tengan el mismo grado de complejidad. De modo que la diferencia entre inteligencia artificial fuerte y débil es una falsa dicotomía.

Además, Searle creó un experimento mental para asentar más aún esa idea, la habitación china, que pasaremos a explicar:

Searle no sabe nada de chino, pero está en una habitación cerrada que se comunica con el exterior por una rendija. A través de ella recibe papeles con mensajes en chino, que no entiende, pero tiene un libro enorme que le explica en su idioma como manipular esos símbolos (con reglas de tipo *if-then-else*) para generar como respuesta otro mensaje adecuado en chino, y entregarlo hacia afuera por la rendija.

Searle simula ser un computador y su enorme libro es el programa con las instrucciones para manipular símbolos, resultado de años de investigaciones en inteligencia artificial fuerte.

La persona china que esté afuera, puede conversar con la habitación intercambiando papelitos pero —y aquí viene la paradoja— nadie en esa habitación comprende chino.

Como resumen de la paradoja, Searle afirma que:

- Los programas de computador son solo sintaxis (lenguajes formales).
- Las mentes humanas tienen semántica.
- La sintaxis no es suficiente para construir la semántica.

Y con ello llega a una contradicción, lo que indica que la habitación china es irrealizable. Es decir, el programa de computador con las instrucciones para manipular símbolos es una entelequia.

Se ha escrito demasiado sobre la habitación china, para apoyarla y para refutarla. Simplemente quiero proporcionar mi punto de vista. El error está en que Searle separa la sintaxis y la semántica como mundos completamente aislados e incommunicados. Pero lo que sabemos actualmente es que la semántica se construye acumulando muchas reglas sintácticas. A veces son reglas deterministas, que siempre se cumplen. A veces son reglas probabilistas con probabilidades condicionadas al disparo de otras reglas. A veces son reglas que pueden coexistir en contradicción. Eso es la semántica. No hay nada más.

En el manejo del lenguaje cada vez se alcanzan nuevos éxitos, como los *chatBot*, los traductores automáticos, los programas que hacen resúmenes de textos, los generadores de poesía y de novelas. Tienen fallos pero cada vez que acumulan nuevas reglas van mejorando. Van construyendo su semántica.

Los humanos somos robots biológicos muy complejos y la semántica también procede de la acumulación de mucha sintaxis. Además, somos propensos a errores, como los computadores. A veces incluso es preferible un error simple de una máquina, fácilmente identificable y corregible, que un error de un humano que entiende, pero mal⁴⁶. Pero con el tiempo las

⁴⁶ Como ejemplo que acaba de ocurrir, un revisor de inglés, experto nativo, me cambió la frase “[...] similar processes can take place on human beings” (“[...] procesos similares pueden ocurrir en se-

máquinas alcanzarán la misma semántica y las mismas posibilidades de error que los humanos.

Para terminar, quiero mencionar que Penrose supone que dentro de la tubulina, una molécula que conforma el citoesqueleto de las neuronas, ocurren procesos mecánicocuánticos que otorgan al cerebro una capacidad de cómputo que viola el teorema de Gödel. Desde que lo formuló sonaba mal, pues no está claro cómo se puede violar un teorema matemático a través de un proceso físico. Y conforme pasa el tiempo ni se encuentran fenómenos especiales dentro de la tubulina⁴⁷, ni tampoco se ha encontrado un solo caso de una máquina cuántica violando el problema de la parada. Para aclarar bien las cosas recordemos que el teorema de Gödel dice que cualquier sistema formal suficientemente complejo, o tiene contradicciones internas o hay verdades que no puede demostrar. Los algoritmos evolutivos y, en general, los algoritmos que usan probabilidades o variables estocásticas, aceptan tener contradicciones internas, de modo que se vuelve posible alcanzar todas las verdades sin violar este teorema.

RESUMEN

Usando autómatas celulares la computación universal es muy fácil de lograr de forma espontánea: se requiere una estructura espacial, localmente conectada, y unos 8 bits de información. Y una vez que se tiene computación completa, se abre el camino para la inteligencia artificial, a la vez que se pueden fabricar simuladores. Hemos visto cómo funciona uno (que puede servir simular ambientes en juegos, circuitos electrónicos, procesos industriales, o nuevos mundos) y hemos resaltado que requiere dos dimensiones temporales.

También hemos visto que la computación universal tiene tres límites fundamentales. Primero vimos el teorema de Gödel junto con el problema de la parada, que desde un cierto punto de vista, más que una limitación lo que indican realmente es que las matemáticas y la computación son creativas.

Después vimos la complejidad computacional, que indica que hay muchos problemas que requieren una cantidad de tiempo enorme para resolverse, por lo que son intratables en la práctica. La forma de saltar esa limitación es emplear aproximaciones o algoritmos probabilistas (como los

res humanos” que tiene un pequeño error sintáctico) por “[...] similar processes can take the place of human beings” (“[...] procesos similares pueden sustituir a los seres humanos” que contiene una grave alteración de significado).

47 Durante la edición de este libro, parece que ya se han encontrado. Pero eso no cambia el resto de la argumentación.

evolutivos), que no tienen garantías de funcionar siempre, pero que pueden dar buenos resultados en muchas ocasiones.

Por último vimos el teorema de *No-Free-Lunch* que dice que para el conjunto de todos los problemas de búsqueda posibles, no hay ningún algoritmo que funcione mejor que otro. Y vimos que este límite se puede superar siempre que no tratemos de abarcar todos los problemas posibles.

PARA SABER MÁS

- Douglas Hofstadter (1979). *Gödel, Escher, Bach: un eterno y grácil bucle*. Barcelona: Tusquets Editores.

A pesar de su antigüedad, sigue siendo un gran libro, de los que se leen una y otra vez, ganando más en conocimiento en cada oportunidad. Y se sigue editando, pues la última versión en inglés es de 1999. Además, se puede leer en español, pues la traducción está muy bien lograda. Hofstadter nos presenta una perspectiva inusual de la inteligencia artificial, mezclando de forma muy amena las preguntas filosóficas con teoremas matemáticos, discusiones de casos, temas de computación, fractales y caos, así como cuentos divertidos que ejemplifican algunos de los tópicos. El hilo conductor del libro es la autorreferencia como generadora de complejidad, que se observa tanto en matemáticas (Gödel) como en dibujo (Escher) y música (Bach). Y en tantos otros campos. Reconozco que este es el principal libro que me inspira a escribir el mío.

REFERENCIAS

Libros, artículos y enlaces web

- Carroll, L. (1995). *Matemática demente*. Barcelona: Tusquets Editores.
- Chaitin, G. J. (1999). *The Unknowable*. Singapore: Springer-Verlag.
- Deutsch, D. (1985). Quantum theory, the Church–Turing principle and the universal quantum computer. *Proceedings of the Royal Society*, 400, pp. 97–117. London. DOI: <https://doi.org/10.1098/rspa.1985.0070>
- Erdős, P. y Renyi, A. (1960). *On the Evolution of Random Graphs*. Mat. Kutato. Int. Kozl. 5, pp. 17–61.
- Forster, M. R. (1999). Notice: No-Free-Lunches for Anyone, Bayesians Included. *11th International Congress of Logic, Methodology and Philosophy of Science*. Cracow, Poland. Recuperado el 27 de agosto de 2017. Disponible en: <https://goo.gl/GQYCo3>
- Garey, M. R. y Johnson, D. S. (1979). *Computers and Intractability. A Guide to the Theory of NP-Completeness*. New York: W. H. Freeman and Company.

- Green, D. G. y Newth, D. (2005). Towards a Theory of Everything? - Grand Challenges in Complexity and Informatics. *Complexity International*, 8.
- Liu, C. L. y Layland, J. (1973). Scheduling Algorithms for Multiprogramming in a Hard Real-Time Environment. *Journal of the ACM*, 20(1), pp. 46-61. DOI: <https://doi.org/10.1145/321738.321743>
- Mackey, P. (2008). *Adventures of a Wetware Hacker*. Recuperado el 12 de junio de 2017. Disponible en: <https://goo.gl/Nqg3eT>
- Mitchel, M. y Crutchfield, J. P. (1995). The Evolution of Emergent Computation. *Proceedings of the National Academy of Science of the United States of America*, 92(23), pp. 10742-10746.
- Nagel, E. y Newman, J. R. (1994). *El teorema de Gödel*. Madrid: Editorial Tecnos.
- Penrose, R. (1991). *La nueva mente del emperador*. Barcelona: Grijalbo-Mondadori.
- _____. (1994). *Las sombras de la mente*. Barcelona: Grijalbo-Mondadori.
- Raichle, M. E. (2010). Two Views of Brain Function. *Trends in Cognitive Sciences*, 14(4), pp. 180-190.
- Searle, J. R. (1997). *El misterio de la conciencia*. Barcelona: Paidós.
- Smullyan, R. (1987). *Forever Undecided. A Puzzle Guide to Gödel*. New York: Knopf.
- Wiki (2017a). *Numeración de Gödel*. Wikipedia. Recuperado el 16 de mayo de 2017. Disponible en: <https://goo.gl/5WdCPd>
- _____. (2017b). *Gödel's incompleteness theorems*. Wikipedia. Recuperado el 22 de noviembre de 2017. Disponible en: <https://goo.gl/C2YDac>
- Wolfram, S. (2002). *A New Kind of Science*. Canadá: Wolfram Media Inc.
- Wolpert, D. H. y Macready, W. G. (1997). No Free Lunch Theorems for Optimization. *IEEE Transactions on Evolutionary Computation*, 1(1), pp. 67-82.

Películas y videos

- Bradbury, P. (2012). *Life in life*. Recuperado el 27 de agosto de 2017. Disponible en: <https://goo.gl/jJhSuo>
- Ikergarcia1996 (2016). *Minecraft - Maquina de Turing | Redstone*. Recuperado el 27 de agosto de 2017. Disponible en: <https://goo.gl/MZKMrP>
- IN6TV (2012). *Máquina de Turing - Lego*. Recuperado el 27 de agosto de 2017. Disponible en: <https://goo.gl/Ac3wd8>

Tesis y trabajos de grado en EVALAB

- Cruz, M. A. (2014). *Búsqueda de solapamiento en clusters, usando técnicas de computación evolutiva*. Cali: Universidad del Valle.
- Vélez, P. A. (2013). *Software para la búsqueda de espacios discretos que se aproximen al espacio continuo euclidiano*. Cali: Universidad del Valle.

VIDA

Life as it could be,

CHRISTOPHER LANGTON

Si la Torre Eiffel representara la edad del mundo, la capa de pintura en el botón del remache de su cúspide representaría la parte que al hombre le corresponde de tal edad; y cualquiera se daría cuenta que la capa de pintura del remache es la razón por la cual se construyó la Torre,

MARK TWAIN

Y llegamos al tema de referencia del título, a base de ir ganando complejidad.

Definir la vida es una tarea fútil y lleva a agrias discusiones (por ejemplo, decidir si los virus están vivos o no, o si las sociedades humanas o todo el planeta son un ser vivo). Lo que se suele hacer es establecer una serie de propiedades que suelen cumplir los seres vivos, pero incluso así, para cada una de ellas hay excepciones. Las que habitualmente mencionan los biólogos son:

- Los seres vivos nacen. Pero hay excepciones como el ácaro *Acarophenax tribolii*. Sus huevos eclosionan en el interior del cuerpo vivo de la madre. Siempre aparecen un macho y varias hembras, que se van alimentando del propio cuerpo de la madre. Al alcanzar su madurez el macho se aparea con todas sus hermanas. Las hembras salen a la luz (nacen), mientras que el macho, habiendo cumplido ya su misión, muere dentro de la madre. Es decir, el macho no llega a nacer (Vélez, 1998, p. 88).
- Los seres vivos crecen. Pero hay excepciones como los virus, que no crecen. Las bacterias prácticamente tampoco. Y también hay contrae-

jemplos como los cristales que nacen y crecen aunque no los consideramos seres vivos.

- Los seres vivos tienen metabolismo. Todos los seres vivos capturan materia y energía del entorno, los procesan para automantenerse, y expulsan lo que no les sirve. Pero el fuego también lo hace y nadie piensa que esté vivo.
- Los seres vivos se reproducen. Esta es sin duda la característica fundamental de la vida. Sin embargo, algunos biólogos no se ponen de acuerdo en ciertos detalles. Por ejemplo, dicen que los virus no están vivos porque no se reproducen por sí mismos; no tienen aparato reproductor, y lo que hacen es manipular los mecanismos celulares de otro organismo para lograr que les reproduzca a ellos. Pero se olvida que los animales superiores ni siquiera pueden sobrevivir por sí mismos sino que necesitan otra serie de animales y plantas que les proporcionen los nutrientes básicos. Los seres vivos no lo están de manera aislada. Están vivos gracias a que pertenecen a una red de seres vivos, con muchas interdependencias entre ellos. Que los virus dependan de otros organismos para reproducirse solo es un caso más de dependencia. Y también hay contraejemplos, como ciertas arcillas que se depositan en lagunas de aguas muy quietas. Las arcillas están formadas por iones de modo que, después de depositarse en el suelo de la laguna, la primera capa que se forme atraerá a los iones complementarios para formar una segunda capa, que a su vez atraerá iones complementarios para formar una tercera capa, y así sucesivamente. Lo interesante es que a veces se desprende la capa de arriba, que cae al suelo de la laguna sirviendo como semilla para hacer crecer otro pequeño montículo de arcilla, que va a tener una distribución iónica similar al “montículo padre” de donde se desprendió. Podemos decir que se ha reproducido, sin estar vivo. El concepto lo leí en Stewart (1998) y tuve la suerte de encontrarlo y fotografiarlo (Figura 158).
- Los seres vivos mueren. La mayoría de los seres superiores sí, con algunas notables excepciones como cierto tipo de medusas (*Turritopsis Nutricula* que, una vez alcanzada su madurez sexual, es capaz de rejuvenecer y puede repetir este ciclo indefinidamente); y algunos peces planos (como el rodaballo de la figura 59), moluscos, crustáceos, reptiles y anfibios, que disminuyen su tasa de mortalidad y aumenta su fecundidad conforme aumenta su edad (Klarsfeld, 2002). Pero además, las células procariotas no suelen morir. Cuando una ameba se reproduce, se parte en dos: la original se convierte en dos amebas hijas, pero eso no quiere decir que nadie haya muerto, pues en el proceso no

queda ningún cadáver. Definitivamente la muerte no es algo obligatorio y responde más a un proceso de optimización (ver el problema 11 de ingenio). Y, por supuesto, que también hay contraejemplos, pues hasta las estrellas mueren, aunque nadie las considera vivas.



Figura 58. Formación de arcillas en una laguna actualmente seca, en una cueva de Bahía, Brasil



Figura 59. Rodaballo en una pescadería (garantizo que no murió de muerte natural)

Obviamente, hay otras características más técnicas, que podrían definirla mejor:

- Disminución local de la entropía.
- Autoorganización.
- Es compleja y crea complejidad.
- Emergen nuevas propiedades.
- Evoluciona.
- Hay impredecibilidad en lo que va a hacer, es decir, goza de cierta libertad.
- Independencia respecto al entorno.

Problema 11: Lo que Henry decidió

En el libro *Genoma* de Matt Ridley (2000) se explica que Henry Ford envió a sus ingenieros para que buscaran todos los coches vendidos en USA y que hicieran un reporte de su estado. Como resultado de ello se dio cuenta de que había muchos en funcionamiento y otros dañados, pero algo común a todos ellos es que el timón estaba perfecto, impecable, como nuevo. Si usted es ingeniero, conteste la siguiente pregunta: ¿qué conclusión sacó Henry Ford de ello? La evolución es un proceso de optimización y hace exactamente lo mismo.

Estas características son más abstractas. Surge la pregunta de si todos los sistemas que las cumplan se pueden considerar vivos. Por ejemplo, si se pudieran diseñar sistemas electrónicos, mecánicos o de *software* que cumplieran con todo ello, ¿se les consideraría vivos? Hay quienes piensan que no, porque la vida está basada en la química del carbono. Estas personas muestran antropocentrismo, que no es nuevo en la historia de la humanidad (somos los únicos, somos los únicos seres pensantes, somos el centro del universo y nuestra química del carbono es la única válida). Se les llama jocosamente *carbonocéntricos*. Por el contrario, hay otras personas que pensamos que sí, pues si la definición de vida es abstracta, la vida como tal también lo es. Y de ahí se ha acuñado el término de vida artificial, que consiste en el estudio de las propiedades de la vida, independientemente del soporte físico empleado. La mejor forma de hacer este estudio es basado en *software* porque es una de las cosas más inmateriales que conocemos.

Problema 12: Campeón en reproducción

Como decíamos, una propiedad de los seres vivos es que se reproducen. ¿Has pensado alguna vez quién es el que se reproduce más rápido en estos tiempos que vivimos?.

VIDA ES INTELIGENCIA

Decimos que un sistema está vivo si muestra cierta independencia respecto al entorno, si puede sobrevivir ante sus cambios y si busca las mejores condiciones para sobrevivir, como más comida, más energía y más posibilidades de reproducción. Un sistema es inteligente si ahonda más en ello, creando herramientas e incluso modificando el entorno para facilitar todavía más su supervivencia, es decir, hacen referencia al mismo conjunto de características. El concepto es idéntico, y la diferencia es solo una cuestión de grado. Nadie negará que para estar vivo se requiere cierta inteligencia.

De modo que si queremos estudiar la inteligencia artificial deberíamos preguntarnos primero por algo más básico, pero que en el fondo es lo mismo: la vida artificial.

Las estructuras que crea un ingeniero son jerárquicas, como los diagramas *UML* en *software* o los de *VHDL* en electrónica: a partir de los elementos básicos disponibles se crean sistemas sencillos que a su vez se usan como ladrillos de sistemas más complejos, y estos a su vez son los ladrillos de otros sistemas todavía más complejos, de forma similar a cómo con ladrillos se construyen paredes, con ellas, casas y a partir de ellas, ciudades. La separación entre niveles es estricta, con lo que se logra disminuir la complejidad del sistema: quien diseña una ciudad no se preocupa por el material y resistencia del ladrillo, porque eso ya fue resuelto por el que diseña una pared. No se preocupa por el amarre con vigas de las paredes, porque eso ya fue resuelto por el que diseña casas. Solo se preocupa por tomar las casas ya terminadas y acoplarlas de manera que se cumplan ciertos requisitos como el acceso por medio de vías (calles, tuberías, cables...) a todas las casas, distancias máximas y detalles similares. De otra manera, el problema sería demasiado complejo para ser abordado. Por ejemplo, si al acercarse dos casas hubiera que cambiar el tipo de ladrillo, o si al seleccionar el material de las vigas hubiera que preguntarse para qué tamaño de ciudad. Romper un problema en partes jerárquicas es la manera ingenieril de evitar la combinatoria de posibilidades, imposibles de analizar una a una en la práctica.

Sin embargo, la vida no trata de minimizar la complejidad de forma tan estricta. Incluso cuando en seres vivos se forman subsistemas (átomos, macromoléculas, células, órganos...), puede haber fuertes interacciones entre niveles distantes: la propagación de un impulso nervioso de extremo a extremo de un ser vivo va a depender de condiciones que se dan a nivel atómico o molecular (concentraciones de Na, K, Mg...). El estado de buen o mal humor de una persona puede depender de concentraciones de macromoléculas neurotransmisoras en su cerebro. Y la reacción adecuada o inadecuada

del sistema inmune de una persona frente a un virus puede depender de su estado de ánimo. Hay interdependencias no jerárquicas.

Esto se debe a que la evolución trabaja como un *bricoleur*⁴⁸ (Francois, 1977) que aprovecha ingenua e ingeniosamente las oportunidades que se presenten, aun cuando ello resulte en diseños complicados. Un ingeniero tiene mucho cuidado en evitar esos diseños porque:

- Es muy difícil analizar matemáticamente el resultado.
- Es muy difícil realizar verificaciones y pruebas sobre sistemas altamente acoplados. Un sistema de dos estados con N partes internas requiere 2^N verificaciones. Mientras que si las N partes podemos separarlas en N/M grupos independientes de M elementos cada uno, entonces solo se requieren $2^{MN/M}$ verificaciones, lo cual es bastante menor. Por ejemplo, para $N=1000$ y $M=10$, entonces $2^N \approx 10^{300}$ mientras que $2^{MN/M} \approx 10^5$.
- Puede llevar a callejones sin salida, donde el diseño no pueda ser modificado buscando mejoras, sin destruir por completo otras características vitales del sistema. Por ejemplo, en biología, el nervio laríngeo apareció en los peces uniendo el nervio vago, que sale del tallo cerebral, con lo que luego sería la laringe. Podría haberlo hecho de dos formas distintas, por encima o por debajo de la vena aorta y, para aquellos peces daba igual. Por azar, lo hizo por debajo. Esa “decisión equivocada” la arrastramos los mamíferos, pues conforme el cuello se estiró, el nervio laríngeo tuvo que hacerlo también, bajando hasta el pecho, pasando por debajo de la aorta y volviendo a subir hacia el cuello. En el caso de los humanos son 15 centímetros de más, mientras que, en el caso de la jirafa, el tonto recorrido del nervio es de 5 metros. Llega un momento donde no se puede estirar más, pues las señales llegarían con mucho retraso.

La vida no sabe nada de esto y, efectivamente, puede meterse en callejones sin salida. Además, la vida no necesita analizar para diseñar, y las pruebas no siguen ningún procedimiento formal sino práctico: simplemente prueba masivamente los nuevos diseños y selecciona evolutivamente a los mejores.

Por cierto: un automóvil antiguo es uno de los peores sistemas diseñados por un ingeniero. La interacción entre las partes es muy alta, lo cual con-

48 Un *bricoleur* es una persona que resuelve problemas con muy poca planeación pero con mucho ingenio, usando de forma oportunista cualquier pieza o herramienta que tenga a mano, así sea de manera distinta para la que fue diseñada. Actualmente son muy populares los almacenes para este tipo de personas, donde venden herramientas y materiales, bajo el lema “hágalo usted mismo”.

lleva a un ahorro en materiales pero a la vez se hace muy difícil predecir la consecuencia de problemas aparentemente triviales, o de poder averiguar que pieza falló, dado un conjunto de síntomas. Por ejemplo, si el motor se apaga, ello también produce un fallo en los frenos, dado que la compresión del motor es la fuerza que actúa sobre ellos. Esto no ocurriría si la fuerza de frenado procediera de un sistema aparte al motor de tracción. Sin embargo, esto a su vez sería más caro. Si dejamos la radio conectada toda la noche, a la mañana siguiente las luces no funcionan y el carro no arranca, debido a que los tres subsistemas comparten la misma batería descargada. Un amortiguador malo no solo produce más vibraciones, sino que también afecta la estabilidad en las curvas e incluso el tiempo de vida de las llantas. Y si se enciende el aire acondicionado se merma la potencia del motor. Etcétera. Es un sistema altamente acoplado, diseñado así para reducir costos. En los automóviles modernos esos problemas tienden a desaparecer porque se hace una mejor ingeniería, desacoplando subsistemas.

VIDA ES REPRODUCCIÓN

A la reproducción también se le puede llamar replicación o autocopiado. Dado que este es el proceso más complejo que interviene en la evolución, vamos a estudiarlo con más detalle. En el capítulo sobre autómatas celulares del libro anterior ya se ha mostrado que se necesitan 2×10^5 bits de complejidad (Figura 60) para lograr la reproducción. Esta es una cota superior, como todas las medidas de complejidad, y es de esperar que se logre reducir usando mejores modelos, pues es un número demasiado alto para alcanzarse en poco tiempo con procesos al azar. Hay que tener en cuenta que von Neumann construyó su modelo de replicación usando un autómata celular de 2 dimensiones y sin utilizar computación completa. El espacio bidimensional tiene demasiadas restricciones que desaparecen en 3 dimensiones (ver el problema 13). Y es muy probable que emerja computación completa, lo que nos da una potencia de trabajo extra. De modo que aunando estas dos cosas, espacios tridimensionales y computación completa, es casi seguro que se logre rebajar esa cota sustancialmente. Sin embargo, conseguir reproducción es realmente difícil y prueba de ello es que en nuestro planeta parece que solo ha ocurrido dos veces: con el ADN y con las arcillas.

Otra forma de explorar la autocopia es partiendo de un sistema suficientemente complejo, como un computador tradicional con lenguaje ensamblador, o una máquina virtual con un lenguaje con menos instrucciones, pero orientadas a la labor de copiado. Dado que la computación completa requiere pocos bits de complejidad, parece razonable esperar que la repro-

ducción sea solo un poco más compleja. Pero no hay estudios sistemáticos sobre el tema. Sí hay algunos particulares como es el caso de *Tierra* (que veremos enseguida) y de *Core Wars*, que sigue la misma idea pero como juego de competición entre programadores. Aquí se trabajaba en ambientes computacionales virtuales, completamente controlados, para evitar que estos programas se propagasen a otros computadores. Pero siguiendo estas ideas surgieron los famosos e infames virus de computador actuales, donde lo principal es que pueden reproducirse saltando de un computador a otro por medio de Internet, memorias extraíbles o similares.

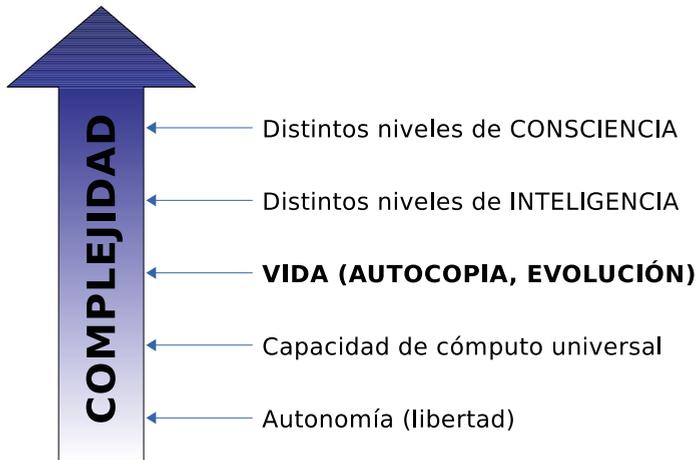


Figura 60. Se requiere 2×10^5 bits de complejidad para lograr la autocopía

Un ejemplo muy interesante de reproducción lo vemos en el robot de 4 cubos iguales (Steele, 2005), del profesor Hod Lipson, que es capaz de sacar una copia de sí mismo. En el video (Skitterbot, 2009) podemos ver cómo lo hace. Mis estudiantes siempre me reclaman que hay trampa, pues parte de piezas complejas ya construidas que se le colocan al lado para que las use. Pero no es tanta trampa si pensamos que nosotros comemos proteínas fabricadas por otros seres vivos. Nadie se alimenta de tierra. Es solo una cuestión de definir cuál es el nivel de partida.

La utilidad de poder hacer copias nadie la pone en duda: los electrones son copia unos de otros⁴⁹. Los ingenieros usan continuamente “copiar y pe-

49 John Wheeler se preguntaba por qué los electrones son todos idénticos y ofreció como posible explicación que se trataba de un solo electrón, yendo y viniendo en el tiempo. Esta explicación fue divulgada por su alumno Richard Feynmann durante la aceptación del Premio Nobel de Física de 1965. La idea es que cuando un electrón choca con un positrón se aniquilan ambos y generan

gar”, no solo los informáticos⁵⁰. Es la operación básica más potente que hay. A diario hay que sacar copias de documentos, de planos, el destornillador que tengo en mi casa existe también en millones de otros hogares y lo mismo se puede decir de cualquier herramienta, *gadget*, libro o película. La humanidad en general prospera gracias a que copiamos⁵¹ e incluso mejoramos (en cierto modo esos son los errores de copia) las ideas e inventos de otros. Podríamos decir que es la humanidad quién es inteligente, y no tanto los humanos individualmente considerados, y de ello se hablará en el capítulo de inteligencia colectiva.

Problema 13: Comunicaciones en 2D

Tenemos 3 emisores de datos {A, B, C} que deben conectarse a 3 receptores de datos {A', B', C'} sin que se haya cortocircuitos entre cables (Figura 61-a). La solución la podemos ver en la figura 61-b. Ahora tenemos el mismo problema pero hemos movido de sitio dos receptores (Figura 61-c). Encontrar la nueva solución no es trivial. ¿Quieres intentarlo?

Figura 61. Cables en 2D

un fotón. Ello también puede interpretarse como un electrón, de carga negativa, que avanza en el tiempo, interactúa con un fotón y rebota hacia atrás en el tiempo con carga contraria (a lo que llamamos positrón). Luego ese positrón interactúa con otro fotón y vuelve a saltar hacia adelante en el tiempo, con carga negativa. Ver figura 62.

- 50 A mis estudiantes siempre les digo que si hay algo ya hecho y que funcione, ¡úsenlo!, dando el debido crédito a los autores. Y que el único ambiente donde no es razonable usar “copiar y pegar” es en el educativo, porque si los profesores lo permitiéramos, no habría forma de saber si los estudiantes han aprendido algo o no.
- 51 Hay un nuevo movimiento de opinión, que suscribo completamente, acerca de lo perjudicial que comienzan a ser las patentes y las protecciones de los derechos de autor. En el siglo XIX eran mecanismos importantes para conseguir financiación y acelerar el progreso de la humanidad. En el siglo XXI están frenando ese progreso; y la financiación puede conseguirse por otras vías, como demuestran los movimientos “open” y “free”: *freeware*, *freehardware*, *open robotics*, etc. Copiar un diseño, una idea, una película, un libro, no para beneficio económico personal, sino para usarlo, aprender, mejorarlo, etc. (dando los debidos créditos al autor) no debe considerarse una actividad delictiva sino la forma natural de funcionamiento de los colectivos.

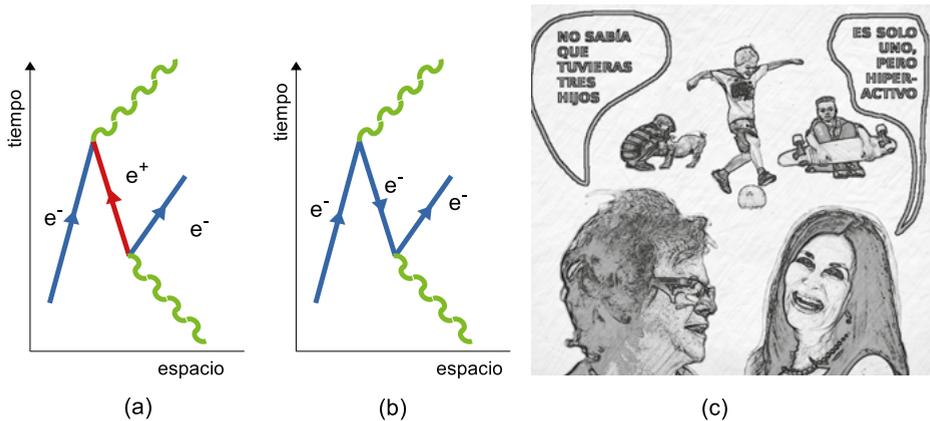


Figura 62. a) Dos electrones y un positrón moviéndose hacia adelante en el tiempo; b) un solo electrón moviéndose hacia adelante y hacia atrás en el tiempo; c) un solo niño

En abstracto la reproducción puede ocurrir principalmente de tres formas:

- El organismo desarrolla una parte de él dedicada a la reproducción. Esa parte especializada es capaz de regenerar todo el organismo. Ejemplos:
 - Una célula eucariota se reproduce dividiéndose en dos, pero el proceso está gobernado principalmente por el núcleo.
 - Un hormiguero saca copias de sus hormigas y eventualmente llegará a dividirse en dos colonias, siendo que el proceso reproductivo está centralizado en la hormiga reina.
 - Un animal o una planta tiene hijos, y el proceso de reproducción ocurre gracias a sus órganos sexuales.
- Cada parte del organismo se reproduce por separado, y los hijos se vuelven a juntar (respondiendo a un mecanismo genético o social) para producir otro organismo. Ejemplos:
 - Las empresas utilizan cualquier persona que pase por allí, para reproducirse generando una sucursal. Solamente se revisan sus funcionalidades (hoja de vida, títulos y experiencia). Al respecto, los estándares de calidad ISO que tratan de hacerse uniformes internacionalmente apuntan a ello: facilitar que un puesto de trabajo sea ocupado por cualquier persona que cumpla ciertos requisitos. Los trabajadores se reproducen y sus hijos formarán parte de otras empresas, por la presión social de tener que buscar trabajo para sobrevivir.
 - El ADN utiliza cualquier molécula que pase por allí, de tipo A, T, G, C para reproducirse. Solo busca la funcionalidad adecuada, en este caso, que sea complementaria de la otra hebra de ADN.

- El animal acuático llamado *Nanomia* es realmente una colonia de varios animales (Smith, 2001, p. 206). Este puede parecer un ejemplo excepcional, pero no lo es tanto. Incluso los seres humanos estamos formados por un porcentaje de células muy alto que no tiene nada que ver con nuestro genoma pero que, si las matamos, morimos nosotros después. Son bacterias que están principalmente en el intestino, pero también en la piel y los vasos sanguíneos, y hay tantas como células humanas (Sender *et ál.*, 2016).
- La maquinaria reproductiva es externa al organismo. El organismo convence o manipula al propietario para que la use para reproducirle a él. Ejemplos:
 - Los virus, que invaden una célula y la manipulan para usar su sistema reproductivo para su propio beneficio.
 - Los memes, que manipulan cerebros humanos para lograr copiarse de unos a otros.
 - Todas las variantes de reproducción por ingeniería genética, fuera del útero o con madre de alquiler.
 - Los objetos de plástico y otros materiales, que se reproducen usando impresoras 3D.
 - En el problema de ingenio “Campeón en reproducción” se muestra otro ejemplo muy sorprendente.

A su vez se puede hacer una taxonomía teniendo en cuenta el papel que juega el grupo respecto al individuo cuando ocurre la reproducción:

- Si el individuo nace dentro del grupo (como las macromoléculas dentro de las células y las personas dentro un país) o fuera y luego se incorpora a él (como los trabajadores de una empresa y los futbolistas en un club).
- Si el individuo puede cambiar de grupo (como los trabajadores de una empresa, las personas de país o los aminoácidos de un cuerpo cuando son devorados por otro) o no puede (como los organelos de una célula y las células de un cuerpo a otro).
- Si el individuo se reproduce a la vez que el grupo (como los organelos en células) o independientemente (como los trabajadores de una empresa).
- Si al morir el grupo mueren los individuos que lo conforman (como los organelos en células y las entidades oficiales de un país) o sobreviven (como los trabajadores de una empresa y los aminoácidos de una proteína).

En un nivel emergente N se da por hecho que existe la infraestructura del nivel anterior $N-1$. Por ejemplo, las empresas cuentan con la existencia de trabajadores, los carnívoros cuentan con la existencia de herbívoros, las plantas cuentan con la existencia de agua y minerales y aire, y las moléculas cuentan con la existencia de átomos.

VIDA ES EVOLUCIÓN

Vemos entonces que la vida, al igual que la realidad, la inteligencia, etc., es un concepto difuso. Aun así, hay una característica muy importante presente en todos los seres vivos: cuando la complejidad de un sistema es tan alta que puede sacar copias de sí mismo (reproducirse) entonces emerge un nuevo fenómeno, la evolución.

La teoría de la evolución fue propuesta por Charles Darwin en 1859 en el ámbito de la biología. Pero lo que seguramente no podría imaginar en aquella época es que la evolución es algo más general: es un algoritmo. Además, es uno de los algoritmos más simples y potentes que hay, aunque como contrapartida, es lento. El que mejor lo entendió fue Lawrence Fogel, quien en 1966 enunció los cuatro requerimientos para que haya evolución, independientemente del sustrato (biológico, químico, mecánico, electrónico, de *software*, económico, de ideas, etc.). Los podemos ver en el recuadro 5.

Recuadro 5: El algoritmo de la evolución

Para que haya evolución se requieren cuatro cosas:

- una población de entes...
- que saquen copias de sí mismos, es decir, que se reproduzcan sexual o asexualmente o de cualquier otra manera tal que los hijos se parezcan mucho a los padres...
- con errores en las copias, es decir, que aunque los hijos sean muy parecidos a los padres, no sean idénticos...
- y que estén sometidos a una presión selectiva. Esto significa que no todos los entes sacan copias de sí mismos en la misma cantidad. Cuanto mejor resuelvan un cierto problema (en biología, sobrevivir y encontrar pareja) más hijos tendrán.

Para Darwin los entes eran los seres vivos. Pero no es necesario restringirnos a la biología. Estos entes realmente pueden ser cualquier cosa: mecanismos, circuitos electrónicos, compuestos químicos, edificios o robots. Y, en computación, estructuras de datos, autómatas o programas.

La parte más difícil del algoritmo es conseguir que los entes tengan capacidad de sacar copias de sí mismos. Actualmente los edificios mencionados no tienen esa capacidad, pero si algún día la tuvieran y se dieran las otras

tres condiciones, la evolución aparecería de inmediato. Es más, sería imparable. No es trivial hacer desaparecer un proceso evolutivo ni frenarlo, y habitualmente es imposible. Por eso tenemos tantos problemas de salud con los virus y las bacterias.

El etólogo Richard Dawkins, en los dos primeros capítulos de su libro *El Gen Egoísta* explica muy bien cómo funciona este algoritmo. Recomiendo la completa lectura de este libro, pues tiene ejemplos sorprendentes e ilustrativos de la evolución, aunque limitados al ámbito biológico. A continuación presentaremos un resumen con lo esencial para comprender el algoritmo.

Para ello comenzaremos con la pregunta que abre ese libro: ¿qué cosas son las que perduran? Si visitamos un planeta y luego retornamos a él unos miles o incluso millones de años después, ¿qué cosas seguirán allí? Obviamente, en primer lugar, las cosas más estables, como montañas y mares. Y en segundo lugar, las cosas que se autorrepliquen, pues si un objeto saca copias de sí mismo lo encontraremos con facilidad innumerables veces (la complejidad que se requiere para que haya autorreplicación la analizamos en el capítulo “Autómatas Celulares” del primer libro). Es decir, conforme aumenta la complejidad de un objeto es menos probable encontrarlo espontáneamente (Figura 63). Piense en lo improbable que debe ser encontrarse un hacha de piedra en la superficie de la Luna. Hay un montón de objetos muy complejos que es prácticamente imposible que existan espontáneamente. Pero después de este “desierto existencial”, aparecen objetos mucho más complejos cuya probabilidad de encontrarlos es prácticamente uno, si se dan las condiciones adecuadas. Son los objetos autorreplicantes, los capaces de sacar copias de sí mismos.

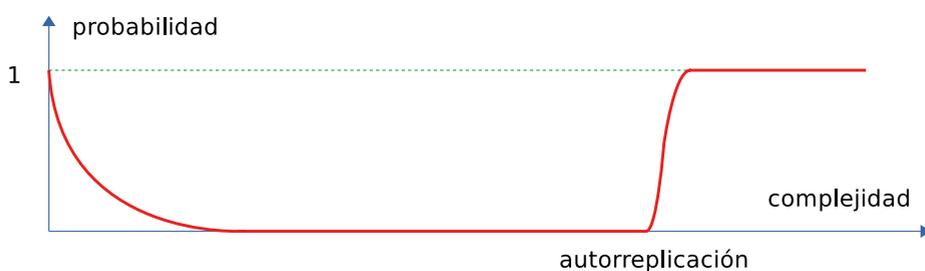


Figura 63. Umbral de autorreplicación

Mientras estamos en esa zona desértica, la combinatoria es el único fenómeno que está explorando posibles estructuras para aumentar la complejidad. Es una búsqueda completamente al azar, sin ningún tipo de guía. Por eso es tan difícil llegar a la autorreplicación. Una vez que se alcanza la autorreplicación, si es que se alcanza, aparece también la evolución que

incluye un bucle de realimentación positiva: cuantos más entes haya que se reproduzcan, más entes habrá. Y un bucle de realimentación negativa: la presión selectiva dificulta la reproducción a los entes menos adaptados al medio, de modo que busca óptimos. Y si te alejas de esos óptimos te va a ir mal. A partir de allí, gracias a estos dos bucles de realimentación, la complejidad se mantiene e incluso crece rápidamente, dando lugar a la inteligencia y a la consciencia.

Todo esto es un resumen rápido de lo que nos espera, y de momento merece la pena situarse en el umbral donde aparece la autorreplicación, para analizarla más despacio y entender las consecuencias. Supongamos entonces que ya tenemos autorreplicación y que, para ponernos en el escenario biológico de Dawkins, existe una molécula capaz de realizar esa tarea en el caldo primitivo en los orígenes de la Tierra. Dawkins discute los mecanismos que pudo emplear esa molécula para lograr la autoduplicación, pero eso no nos interesa aquí y, de hecho, han aparecido otras propuestas distintas⁵².

Como decíamos, partimos de una molécula A capaz de sacar copias de sí misma usando los elementos químicos que tiene alrededor (Figura 64-a). Inevitablemente se producirán algunos errores en las copias (Figura 64-b), por ejemplo porque para una reacción química se requiere un elemento que está escaseando pero a cambio existe otro con propiedades parecidas (esto ocurre con los elementos que están en la misma columna de la tabla periódica, como el sodio y el potasio, que son intercambiables en muchas reacciones, aunque el resultado final no es exactamente igual). Por ello en algún momento aparecerán variantes de esa molécula algunas de las cuales seguirán teniendo la capacidad de autocopiado, como A_1 , A_2 , etc.

Es obvio que el autocopiado es un proceso de realimentación positiva que produce un crecimiento exponencial. Y las exponenciales tienen varios parámetros. Por ejemplo, la molécula A saca dos copias de sí misma cada cierto tiempo, mientras que la molécula A_1 solo saca una copia de sí misma, aunque lo hace más rápidamente. La molécula A_2 es más lenta y la A_3 es más prolífica.

52 Ver Patel, Percivalle, Ritson, Duffy y Sutherland (2015), donde se explica que el ácido cianhídrico puede generar los precursores de los ribonucleótidos, los ácidos y los lípidos necesarios para llegar al “mundo del ARN”. A su vez, el ácido cianhídrico procede de la desestabilización de las órbitas de los asteroides por las interacciones de Júpiter y Saturno, que produjeron el último gran bombardeo sobre la Tierra. El carbono de los meteoritos reaccionó con el nitrógeno de la Tierra para generar el ácido cianhídrico.

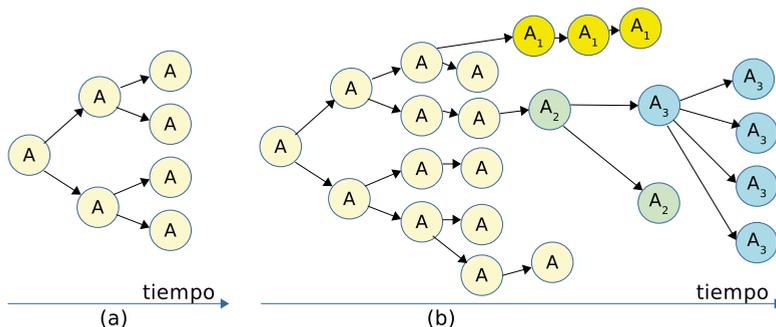


Figura 64. Autorreplicación: (a) exacta y (b) con mutaciones

Si hay moléculas autorreplicantes similares a estas en un charco de agua con minerales, hay varios factores que influyen en lo que voy a encontrar cuando regrese unos años después:

- **La longevidad.** Cuanto más estable sea la molécula, más tiempo tendrá para sacar copias de sí misma.
- **Lo prolífica que sea.** Es decir, la cantidad de copias que saque por unidad de tiempo.
- **Lo exacta que sea.** Si las copias son imperfectas, las originales serán más difíciles de encontrar.
- **Lo agresiva que sea.** Conforme los minerales del charco se vayan agotando, quizás alguna de esas moléculas descubra, por mutación, que las otras moléculas son también una fuente de esos minerales. De este modo aparecerán las primeras moléculas protocarnívoras, capaces de descomponer otras para reutilizar sus componentes. Ello tendría una doble ventaja: por un lado, conseguir minerales para sí misma; por otro, disminuir la competencia de las otras moléculas.
- **Las protecciones que desarrolle.** Las moléculas capaces de protegerse de esas protocarnívoras tendrán ventajas frente a las que no puedan hacerlo. Ello puede traducirse en crear barreras como, por ejemplo, membranas celulares.

Esto es una guerra que, en evolución y en teoría de juegos, se llama “carrera de armamentos” o “carrera de la Reina Roja”⁵³, cuyo único objetivo es sacar copias de sí mismo. Con el paso de eones se desarrollaron estrategias muy sofisticadas como la encapsulación de estas moléculas re-

53 En “Alicia a través del espejo”, Lewis Carroll en una parte del cuento Alicia y la Reina Roja de la baraja recorren un camino que se mueve hacia atrás cada vez que ellas da un paso hacia adelante, por lo que tienen que correr mucho para poder mantenerse en el mismo sitio.

plicadoras (que hoy llamamos genes) dentro de torpes robots controlados por ellas. Dawkins los llama “máquinas de supervivencia” y nos llena de asombro cuando comprendemos que son los vegetales, animales e incluso nosotros, los humanos, y que siguen manteniendo el mismo y único fin: su replicación. El ambiente cada vez más complejo ha hecho que surjan, por mutaciones, características nuevas como cerebros que se comportan con inteligencia y con consciencia, que ayudan a estos robots a tomar decisiones rápidas orientadas a la supervivencia y reproducción de los genes que los fabricaron. Pero ello tiene un costo: la inteligencia va asociada a un cierto grado de libertad, por lo que los cerebros de las máquinas de supervivencia no tienen por qué seguir los dictados de los genes, o no siempre. La evolución de los genes es lenta y no pueden controlar por completo las decisiones de los cerebros, que son muchísimo más rápidas. Los genes van perdiendo libertad mientras que la van ganando los cerebros⁵⁴.

Además los cerebros, usando herramientas físicas y culturales, cumplen rápidamente su misión de supervivencia y de óptimo de reproducción (no tiene sentido reproducirse más allá del óptimo, tanto que las crías carezcan de alimento), por lo que les queda tiempo libre en el que siguen pensando y haciendo cosas que habitualmente son neutrales a la supervivencia y reproducción. Pueden pintar, cantar, resolver acertijos matemáticos y otras cosas raras. Eventualmente muchas de esas cosas también ayudan a sacar copias de sí mismos, pero a largo plazo y como resultado inesperado, pues los genes solo controlan lo inmediato.

La longevidad genera un crecimiento lineal de copias, mientras que ser más prolífico implica un crecimiento exponencial, de modo que cuando un animal tiene que decidir apostar más o menos recursos en una u otra cosa, la evolución presiona hacia tener más hijos. Esa es otra de las razones por las que la diferenciación sexual también lleva aparejado el envejecimiento y la muerte: un crecimiento exponencial siempre gana a uno lineal. Al respecto se puede ver también el ya presentado problema 11.

Muchas veces se ha malinterpretado a Dawkins diciendo que su exposición de lo que es la evolución a nivel genético nos lleva a un determinismo en el comportamiento. Dawkins no dice esto, y si hay libertad o no en nuestras acciones no depende de nuestros genes. Al menos en animales supe-

54 La libertad inevitablemente se irá concentrando en un solo nivel. Entendamos la razón con un ejemplo: si vas conduciendo una bicicleta, tú tienes libertad para decidir hacia dónde quieres ir, y para ello manipulas los controles de la bici (pedales, freno...) que obedece ciegamente. Pero si la bici tuviera libertad (por ejemplo, decidiendo caprichosamente cuando quiere girar a la izquierda) entonces tú ya no la tienes.

riores. El tema ya se mencionó en el capítulo “Libertad”, pero lo que aquí queremos aclarar es que los genes construyen los cuerpos y sus cerebros de manera más o menos determinista (aunque el ambiente también influye mucho en ambas cosas). Sin embargo, lo que los cerebros decidan no está determinado por los genes, salvo en casos de enfermedades muy graves. Entre otras cosas, en el genoma humano no hay información suficiente para decidir dónde van a estar las neuronas cerebrales y con qué otras neuronas van a tener conexiones. En el comportamiento de un individuo influyen sus genes (especialmente en lo que tiene que ver con instintos de supervivencia y reproducción), las hormonas (que a su vez dependen de los genes, pero sobre todo dependen de las interacciones sociales, el ambiente, la comida, la contaminación del entorno...), la sociedad (los padres, los pares, los educadores...) y hay una parte que no se sabe de dónde procede, probablemente de algún acontecimiento aislado y significativo en la infancia, o simplemente el azar, por lo que queda algo de margen para actuar en libertad.

En *El gen egoísta* Dawkins prosigue argumentando que en un ambiente altamente competitivo es inevitable que los genes lo sean, y que las máquinas de supervivencia que generan también sean así. Para ello primero define “altruismo” como la actitud de un individuo para ayudar a otro incluso a costa de su propia supervivencia, así sea en pequeña cantidad o probabilísticamente. En este sentido, un animal que arriesgue su vida para salvar a otro, es altruista. Y una persona rica que regala comida a otra no lo es, puesto que con esa acción no se pone en riesgo ni disminuye su probabilidad de supervivencia. El libro continúa mostrando muchos ejemplos de egoísmo entre los animales, pues eso es lo que cabe esperar, que sigan el dictado de sus genes. Pero también muestra casos de altruismo y advierte que, dado que son una anomalía no predicha por la teoría, hay que analizarlos muy bien para entender de qué se trata. Al final resulta que todas las acciones son solo aparentemente altruistas, son un egoísmo disfrazado.

Dawkins argumenta certeramente que es imposible que exista verdadero altruismo dirigido por los genes. Si tenemos un organismo con un gen que controla el comportamiento, de dos alelos, altruista y egoísta, el organismo que porte el gen altruista va a ayudar a la supervivencia o reproducción de los demás organismos, a expensas de él mismo, de modo que el organismo que porte la variante egoísta se va a encontrar rápidamente en mayor proporción en la población.



Fuente: CC 2.0, Marty Stone (Flickr). Disponible en: <https://goo.gl/bjSe9a>

Personaje 6

Richard Dawkins (1941-)

Richard Dawkins es un biólogo evolutivo, etólogo y un excelente divulgador científico. En su primer libro, *El gen egoísta*, definió lo que sería toda su carrera como divulgador. Ahí proponía una nueva forma de ver la evolución, centrada en los genes. Mucha gente no ha leído el libro y se queda con la falsa impresión de que los genes deben ser malos porque son egoístas, o algo así. No seas tú uno de ellos. Muy por el contrario, el altruismo en los individuos sofisticados emerge como consecuencia del egoísmo de sus genes. En este libro crea el concepto de *meme*, como unidad replicante de información, que se ve ahora tanto en Internet con significados menos claros. Otros de sus libros son *El fenotipo extendido*, *El relojero ciego*, *Escalando el monte improbable*, *Evolución: el mayor espectáculo sobre la Tierra* y *La magia de la realidad*. ¿Cómo sabemos lo que realmente es verdad? También es un firme defensor del método científico que se opone a las religiones, los esoterismos, las medicinas alternativas y las pseudociencias. Es uno de los más conocidos *brights*, o “iluminados, brillantes”, que tienen una concepción naturalista del mundo en contraposición a la supernatural (creacionismo). Ver <https://www.the-brights.net/>

Lo único que puede existir es altruismo para los organismos que compartan muchos genes, es decir, dentro de las familias⁵⁵, como cuando los padres protegen a los hijos. El ejemplo más obvio es el de muchos tipos de hormigas donde, por la forma de reproducirse, las hermanas tienen casi todos los genes idénticos. Por ello a una hormiga le da igual reproducirse ella misma o dejar que la reina (la madre de todas) lo haga, pues genéticamente hablando, el resultado va a ser el mismo. Esta es la razón por la que las hormigas no se reproducen individualmente sino que todas ayudan a que la reina lo haga. La reina es algo así como el órgano sexual del hormiguero. De la misma manera las hormigas están dispuestas a dar la vida defendiendo el hormiguero, porque los genes que sobrevivan serán los mismos suyos.

Sin embargo, el altruismo también puede existir debido a otras razones distintas a los genes, por ejemplo culturales que enseguida pasaremos a ver.

55 Como chiste, el biólogo Haldane decía que daría la vida por 2 hermanos u 8 primos (porque en promedio no se perderían genes).

La evolución dirigida por los genes no es solo un modelo teórico⁵⁶. En el 2009 Dawkins, quien además de ser un excelente científico también es un prolífico escritor, publicó el libro *Evolución: el mayor espectáculo sobre la Tierra* donde muestra con toda certeza que la evolución es un hecho, para que los escépticos pierdan toda capacidad de argumentación en contra. Aparte de miles de evidencias indirectas, también hay experimentos directos que se realizan con moscas o con bacterias, aprovechando que su ciclo de vida es más corto, con lo que es posible obtener cientos de generaciones en cuestión de semanas.

Uno de los experimentos más interesantes es el del bacteriólogo Richard Lenski y su equipo, que comenzó en su laboratorio de la universidad del Estado de Michigan y sigue en marcha actualmente, consistente en un cultivo de bacterias *E. coli* repartidas en doce matraces aislados, y alimentadas por una pequeña cantidad de azúcar, citrato y otros químicos, de los cuales solo el azúcar es alimento para ellas. Este azúcar les dura poco, apenas les da tiempo a reproducirse varias veces y al final del día pasan hambre. Cada 24 horas se extrae una muestra de las bacterias que sobrevivieron y se reintroducen en otros frascos, dándoles la misma dosis de alimento. El frasco original se congela para poder analizarlo posteriormente. De esta manera mantienen doce líneas evolutivas simultáneamente que, al día de hoy han producido alrededor de 50 000 generaciones de bacterias. Inevitablemente aparecen mutaciones, algunas de las cuales usan el azúcar de forma más eficiente, con lo cual terminan imponiéndose numéricamente a las demás. Eso hace que las bacterias aumenten progresivamente de tamaño, y como las mutaciones dependen del azar, cada una de las doce líneas evolutivas siguió un camino distinto. Sin embargo, en la generación 30 000 apareció algo radicalmente nuevo en una de las doce líneas: las bacterias habían aprendido a usar el citrato como comida. Investigando cómo lo lograron, por qué demoraron tanto, y por qué solo una de las doce líneas lo consiguió, se dieron cuenta que en la generación 20 000 apareció una mutación aparentemente neutral, pero que con la mu-

56 Muchas veces se dice despectivamente que la evolución es solo una teoría. Efectivamente, es la “teoría de la evolución”, tan confirmada como la “teoría de la relatividad”, la “teoría cuántica”, la “teoría electromagnética”, etc. Quizás la confusión prospera porque en el lenguaje común, “teoría” significa algo que debe comprobarse. Mientras que en lenguaje científico una “teoría” es un cuerpo de conocimiento sin contradicciones internas, que ha superado muchas pruebas y que es ampliamente aceptado por la comunidad científica. Cuando todavía no ha superado pruebas se le llama “hipótesis” o “conjetura”. Un cuerpo de conocimiento maduro alcanza el estatus de “teoría”, que es el máximo nivel que puede alcanzar. Quizás otra fuente de confusión es que a las primeras teorías se las llamaba de forma arrogante “leyes” (ley de la gravedad de Newton o ley de Snell) pero realmente eran “teorías” (teoría de la gravedad de Newton) o simples fórmulas más o menos exactas (fórmula de la difracción de Snell).

tación de la generación 30 000 combinadas, lograban el nuevo fenómeno. O sea, descubrieron que los cambios suaves se logran con mutaciones aisladas, y los cambios bruscos con mutaciones dobles, cuya probabilidad de ocurrir es bastante más baja. De modo que es un experimento bonito que muestra la evolución en marcha, pero además reconcilia las dos versiones de la evolución: continua (Darwin y Dawkins) y saltacionista (Jay Gould y Niles Eldredge), a la vez que lo hace con la microevolución (optimización progresiva de características) y la macroevolución (aparición de nuevas especies).

Más recientemente, en 2016 el profesor Michael Baym y sus colaboradores realizaron otro experimento con bacterias en una placa de Petri gigante dividida en 9 zonas. En las dos zonas extremas (1 y 9, de la figura 65) no había nada inicialmente. En las dos zonas siguientes (2 y 8) se dispuso un antibiótico en una dosis muy leve. En las dos siguientes (3 y 7) pusieron el mismo antibiótico pero a una concentración 10 veces mayor, en las dos siguientes (4 y 6) era 100 veces mayor y en la central (5) era 1000 veces mayor. Luego se colocó una cepa de bacterias en los extremos donde crecieron sin problemas hasta tropezar con el antibiótico (zonas 2 y 8) que les impidió seguir creciendo. Pero al cabo de algún tiempo, una mutación afortunada logró producir una nueva bacteria resistente al antibiótico, y ello ocurrió en ambas zonas. Las bacterias invadieron las zonas 2 y 8 pero quedaron frenadas al tocar el borde de las zonas 3 y 7. Y volvió a ocurrir lo mismo: en poco tiempo aparecieron mutaciones que aumentaron la resistencia a esa concentración del antibiótico. Al cabo de apenas 10 días, también lograron conquistar la zona central. Los profesores usaron este experimento para prevenirnos contra el mal uso de los antibióticos: nunca hay que usarlos alegremente para cualquier enfermedad (por ejemplo, es inútil usarlos contra virus), siempre hay que tratar de usar el antibiótico más fuerte posible para el tipo de bacteria de que se trate (y por ello primero hay que identificar el tipo de bacteria) y siempre hay que tomar toda la dosis que nos prescriba el médico (es un grave error dejar de tomarla en el momento en que nos sintamos bien). Cualquiera de estas cosas que hagamos mal va a producir cepas de bacterias cada vez más resistentes⁵⁷. De hecho ya se habla de superbacterias para las que ningún antibiótico resulta eficaz. Es recomendable ver también el video (Pesheva, 2016) para tener una idea clara de cómo se produce la evolución. Para vencer los antibióticos las bacterias encontraron muchas mutaciones y allí se puede ver el árbol evolutivo. Es interesante también que

57 Y ello no está exento de controversias pues parece que los propios médicos recetan mayores dosis de las debidas, creando el problema de que también se matan las bacterias amigas (Llewelyn, 2017) dejando el cuerpo más débil frente a futuras infecciones.

las bacterias más resistentes no son necesariamente las primeras en llegar al centro. Al centro llegan las más rápidas, y son seguidas por otras cepas a las que ya se les ha despejado el camino.

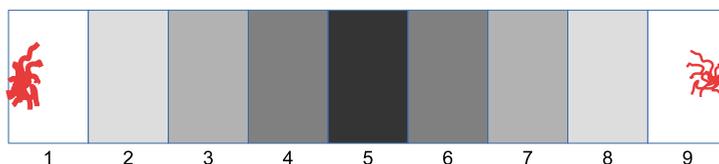


Figura 65. Experimento con bacterias y niveles progresivos de antibiótico

La evolución no suele ocurrir sola, sino que habría que hablar de coevolución: en este caso, las medicinas evolucionan (artificialmente) para ser cada vez más eficaces en matar organismos patógenos, pero estos a su vez evolucionan para ser resistentes a dichas medicinas⁵⁸.

No es difícil ver la evolución en situaciones cotidianas, si uno sabe qué buscar. Hace muchos años me cambiaron a una oficina un tanto inhóspita, con muchos mosquitos. Dados los problemas de salud que estos animales representan, comencé a eliminarlos aplastándolos con las manos. Fue muy fácil porque eran lentos y maté muchos, prácticamente todos. Prácticamente. ¿Prácticamente? Alguno se me escapó porque al cabo de unos días volví a encontrarme con muchos mosquitos, pero más rápidos. Volví a realizar la misma operación, pero para mi desgracia ellos volvieron a aparecer más adelante, y ahora, además, ya no volaban en línea recta, sino que realizaban al azar un rizo que volvía su vuelo impredecible y dificultaba su captura. Cuando me rendí, tenía la oficina llena de mosquitos veloces realizando vuelo acrobático. En este ejemplo tenemos una población de mosquitos que se reproducen. La variabilidad (mosquitos lentos, rápidos, con vuelo rectilíneo o en rizo) ya debería estar allí, en la población inicial, pues no creo que en tan poco tiempo pudieran darse mutaciones tan afortunadas. Y yo era la

58 No hay solución a este problema, pero sí puede retardarse poniendo en práctica las precauciones mencionadas. También podría hacerse una carrera de armamentos *in vitro* entre un grupo de patógenos enfrentados a sucesivas versiones cada vez más refinadas de medicamentos. Después de, digamos, cinco generaciones de patógenos distintos se podrían matar escrupulosamente todos (para que ninguno salga al mundo exterior) y vacunar a la gente con el cóctel de medicinas producidas en todas las generaciones. Un patógeno que ataque *in vivo* a una persona así vacunada moriría; si es capaz de producir una mutación afortunada, también moriría (porque *in vitro* ya ocurrió lo mismo); en el improbable caso de que logre producir una mutación doble también moriría por la misma razón. La probabilidad de que se produzca una mutación quintuple afortunada es bajísima (pero no cero), por lo que la humanidad quedará a salvo de ese patógeno durante muchísimo tiempo, pero no para siempre, y el día que ello ocurra el patógeno será superpoderoso.

presión selectiva. Mataba a los más lentos y predecibles, con lo cual permití que se reprodujeran los más rápidos e impredecibles.

Muchos años después, en el pequeño jardín de mi casa sembramos césped que quedó muy bonito, pero aparecieron malezas. Eran plantas con tallo alto y delgado y llenas de semillas. Lo que hice fue arrancarlas una a una. En semanas sucesivas aparecieron más, posiblemente porque había muchas semillas en el suelo, pero las fui arrancando metódicamente. Al cabo de un tiempo pensé que ya no había ninguna más y pasaron los días creyendo yo que era así. Sin embargo, un día al agacharme en el suelo me di cuenta que estaba completamente equivocado: había mucha maleza, como siempre, pero el tallo crecía casi paralelo al suelo con lo cual, desde lo alto, permanecía prácticamente invisible. De nuevo la presión selectiva era yo, eliminando las malezas que resaltaban porque crecían verticalmente, y dejando que se reprodujeran las que lo hacían horizontalmente. Es muy difícil detener un proceso evolutivo.

Uno de los trabajos pioneros en vida artificial se debe a Tom Ray. Este biólogo se estaba aburriendo en las selvas de Costa Rica tratando de ver en vivo y en directo la evolución, pero como habitualmente es demasiado lenta para detectarla así, decidió crear una simulación en el computador, a la que llamó "Tierra". Para ello definió lo que hoy llamaríamos una máquina virtual dotada de operaciones muy básicas y, usándolas, creó un programa que averiguaba su propio punto de inicio y de final, sacaba una copia en memoria de todo lo que hubiera entre medias, creaba con ello un nuevo hilo y lo activaba. Para evitar que la memoria se llenase, de vez en cuando mataba al azar unos cuantos hilos. Y también introdujo mutaciones al azar de vez en cuando sobre los hilos que estaban funcionando. Con lo primero tenía una población reproduciéndose, con lo segundo una presión selectiva y con lo tercero variaciones en las copias. Es decir, programó todos los ingredientes necesarios para que surgiera la evolución. Y eso es lo que ocurrió pues con asombro vio que en poco tiempo aparecieron espontáneamente muchos otros programas reproduciéndose. La sorpresa mayor vino cuando se dio cuenta que había algunos programas más cortos que el suyo, que se suponía que era el más corto posible. Al investigar su funcionamiento, vio que no tenían toda la funcionalidad de la autocopia sino que hacían uso de trozos de código de otros programas. Es decir, eran el equivalente a parásitos, y por ello lograban ser más cortos que los demás. De forma similar encontró casos de hiperparasitismo y de simbiosis. Esta es una prueba contundente de que la evolución funciona en un caso general y abstracto, como es el *software*, por lo cual es razonable esperar que también lo haga en un caso más específico, como el biológico.

La evolución es una realimentación positiva fuerte y difícil de detener una vez que se ha puesto en marcha. Al respecto, en el 2009 Rafael Moreno realizó su tesis de maestría en EVALAB creando una simulación de un sistema ecológico con predadores y presas moviéndose por una cuadrícula. Había también energía solar que llegaba periódicamente para hacer crecer las plantas en la base de la cadena alimenticia. Era un sistema abierto, pues le entraba energía del exterior, y cada ente tenía capacidad de reproducirse si se alimentaba lo suficiente. Lo que vimos en este sistema es que era muy difícil destruir la vida en él. Había catástrofes y grandes extinciones de vez en cuando, pero la vida siempre se recuperaba. La única forma que encontramos de acabar con toda la complejidad fue apagar el sol artificial durante bastante tiempo. Esto nos dice que la vida debe ser una propiedad muy común en todo el universo. Es bonito ver que en los últimos 15 años la ciencia ha pasado de rechazar esa posibilidad a aceptarla abiertamente, seguramente como consecuencia de que se han descubierto muchos exoplanetas, y varios de ellos con condiciones adecuadas para una vida similar a la terrestre. La NASA acaba de anunciar que la vida puede ser posible en Encélado (satélite de Saturno) además de en Europa (satélite de Júpiter). Incluso en mundos con tan poca temperatura como Plutón la NASA ha encontrado sistemas geológicos complejos. No se puede descartar la vida allí, como tampoco se puede descartar en Venus o incluso Mercurio, a pesar de que las condiciones son adversas, pues simultáneamente se han encontrado en la Tierra variantes de la vida un poco distintas a las tradicionales. No es obligatorio que la vida sea idéntica a la terrestre y podría estar basada en otros elementos químicos. Incluso aquí en la Tierra se han encontrado arqueas a muchos kilómetros de profundidad bajo el suelo, que no necesitan agua ni luz solar para vivir, sino que en su lugar usan monóxido de carbono e hidrógeno (Baker *et ál.*, 2016); también se han encontrado bacterias extremófilas en volcanes submarinos, a altas temperaturas y viviendo con sustancias químicas poco habituales, como la última encontrada en las Islas Canarias por Danovaro *et ál.* (2017); en ambientes muy fríos y salados, sin oxígeno (Murray *et ál.*, 2012); en un lago de California se encontró una bacteria que es resistente al venenoso arsénico, e incluso lo utiliza para sus propios fines (Erb *et ál.*, 2012); además, se sabe desde hace tiempo que muchos hongos, bacterias y arqueas pueden subir a la parte alta de la atmósfera, soportando bajas temperaturas, falta de oxígeno y fuerte radiación ultravioleta (Fulton *et ál.*, 1965), pero es que algunas bacterias aguantan bien en el espacio y ya se han colado en naves de la NASA, regresando vivos (Satomi *et ál.*, 2006) y es posible que hayan llegado así a Marte. La evolución logra que la vida sea muy flexible y adaptable a diversos ambientes.

Hay modelos teóricos de sistemas evolutivos prebióticos, donde por ejemplo se muestran pasos intermedios antes de llegar al ARN y ADN, como en Goldford *et ál.* (2017), donde se propone una química precursora de la vida que ni siquiera necesita fosfatos, que hoy día son imprescindibles.

Pero la evolución no se encuentra únicamente en biología. Dawkins propone otro sistema evolutivo donde la unidad de información son los memes. Un meme es simplemente una idea. Los memes luchan por reproducirse pero no pueden hacerlo por ellos mismos sino que tienen que seducir primero a un cerebro humano para que lo haga, contando la idea a otros humanos, escribiéndola en un libro o enviándola por *e-mail*. Vemos que los memes cumplen las cuatro condiciones para que haya evolución del recuadro 5. Por ejemplo, la ciencia es un proceso evolutivo porque:

- Hay una colección de memes, que son las ideas, conjeturas, hipótesis y teorías científicas.
- Que se reproducen por medio de libros, revistas y congresos de ciencia, usando como herramienta para ello el cerebro de los científicos.
- La reproducción no siempre es exacta porque hay errores experimentales (por ejemplo, los radioastrónomos de Bell Labs que descubrieron la radiación de fondo del universo, estaban buscando caca de pájaro⁵⁹), o los científicos, ingenieros y artistas proponen mejoras o combinan varias ideas para formar una nueva, como cuando en el antiguo Egipto solían construir tumbas en forma de edificio rectangular con los lados inclinados, llamadas *mastabas*; el emperador Zoser pidió a su arquitecto colocar una *mastaba* encima de otra y, repitiendo el proceso varias veces levantó algo nuevo, que hoy se conoce como pirámide.
- Y los memes están sometidos a una presión selectiva, dado que los científicos ponen a prueba las ideas en los laboratorios y ven cuál es la que se adecúa más a la realidad. Esa es la que va a ser más reproducida en revistas y congresos.

59 Al respecto, recomiendo leer *Eureka y Euforias* de Gratzner, un libro muy divertido y enriquecedor. Muestra muchísimos descubrimientos científicos y de ingeniería realizados gracias a la suerte. En este sentido se puede uno preguntar por qué se orientan tan mal los proyectos de investigación, con metas y fechas rígidas, burocratizando todo el proceso, cuando la suerte es un factor tan determinante. La investigación japonesa para fabricar la “quinta generación de computadores” fracasó por esas razones. Solo se puede hacer investigación predecible en resultados a un nivel muy elemental, por ejemplo, voy a sumar 2 más 2 a ver cuánto sale, y dedicaré tantos días y tanto dinero. Estoy seguro de que puedo cumplir los objetivos, pero a eso no deberíamos llamarlo “investigación”, sino “desarrollo”. La investigación requiere cronogramas abiertos, y la posibilidad honorable de fracasar o de que salgan resultados distintos a los objetivos proyectados.

Las ingenierías también son un proceso evolutivo donde se combinan ideas para crear otras nuevas (teléfono + cámara + computador = *smartphone*) y la presión selectiva son los gustos de los compradores. Lo que más se venda se va a reproducir más también. Incluso las ideas políticas o la moda son también sistemas evolutivos. En el caso de la moda, la presión selectiva es la aceptación de los demás frente a mi manera de vestir o de peinarme.

Volviendo al algoritmo evolutivo, otro investigador que ha estudiado de forma abstracta cómo pudo ser este proceso de autoduplicación es Stuart Kaufmann en lo que llama reacciones autocatalíticas⁶⁰. Las conclusiones son muy bonitas porque, por un lado, muestran que es inevitable que ocurra la autocatálisis cuando tenemos muchos compuestos químicos y muchas reacciones entre ellos. Por otro lado, sugiere una relación entre el número de estructuras de alto nivel que se forman (N_c , número de células distintas en un organismo biológico) y el número de reacciones autocatalíticas existentes (N_G , número de genes en ese organismo) que es aproximadamente

$$N_c \approx \sqrt{N_G} \quad \text{Ec. 34}$$

Y, como un tercer aporte, ha servido como inspiración para el desarrollo de químicas artificiales, que son básicamente gramáticas computacionales con reglas de producción. El conjunto de átomos artificiales (que son las proposiciones de la gramática) así como las reacciones químicas entre ellos (que son las reglas de producción) se generan al azar. Esto es muy importante, pues si uno quiere estudiar la emergencia de propiedades desde una concepción naturalista del mundo, es decir, no teleológica, es imprescindible que las condiciones iniciales y las reglas no sean muy específicas, no estén diseñadas *ad hoc* para conseguir cierto resultado. Por ello las químicas artificiales resultan muy interesantes, al no disponer de los conocidos átomos hidrógeno, helio, litio, berilio, etc, sino de otros abstractos que se fabrican al azar. Las conclusiones que se sacan de estos sistemas son más generales. Y, a pesar de la aleatoriedad, efectivamente salen conclusiones. En este sentido, en el 2007, Speroni y Diitrich nos presentaron un trabajo muy interesante: en el computador crearon una química artificial aleatoria, dejando que se desarrollaran reacciones en el tiempo. El espacio era digital (una matriz de celdas) y solo ocurrían reacciones entre los productos químicos que estaban en la misma celda. Al observar el sistema a diferentes escalas apareció una superposición de muchos compuestos químicos nuevos y en distintas proporciones en función de la escala. Lo que querían saber es si había emer-

60 Catálisis: una reacción química que se acelera en presencia de un cierto material, que pasa a llamarse catalizador. Autocatálisis: cuando dicha reacción química produce su propio catalizador.

gencia de estructuras y a que escala ocurrían, para lo que realizaron dos simulaciones: una normalmente y la otra agitando, es decir, intercambiando al azar celdas de vez en cuando. Ellos observaron que había diferencias entre las dos simulaciones, de donde dedujeron que había estructuras que la agitación destruyó. Este trabajo es muy bonito e ingenioso.

Otro tema polémico es la evolución humana. Se dice equivocadamente que la evolución nos está empujando a los humanos a tener el cerebro cada vez más grande para ser más inteligentes (y ahora, curiosamente, también aparece la otra opción, cada vez más pequeño) o a que solo tendremos dos dedos porque eso es lo que se necesita para manejar el *mouse* del computador. Se dicen muchas tonterías. Sin duda que esos rasgos pueden aparecer debido a mutaciones al azar, pero ninguno de ellos garantiza tener más hijos, de modo que los genes causantes no van a inundar nuestras poblaciones. La verdad es que tener hoy más o menos hijos no depende de ningún rasgo esencial. Ser más atractivo para encontrar más parejas depende de modas, que pueden cambiar volublemente con el tiempo. E incluso así, los métodos anticonceptivos ampliamente difundidos hace que ello no se traduzca a tener más hijos.

Se dice frecuentemente que la evolución se ha detenido en los humanos, debido al avance de la ciencia y la medicina. Así, personas que en un entorno aislado podrían fácilmente morir (debido a enfermedades o incapacidades físicas o cognitivas), dentro de nuestra sociedad perviven y además pueden perfectamente tener hijos. De modo que ya no hay presiones evolutivas.

Esto tampoco es cierto. Hay estudios (Rosling, 2010) que indican que el número de hijos de un grupo social depende inversamente del grado de riqueza, de la educación, del acceso a servicios sociales (como la salud y las pensiones de jubilación) y, por supuesto, del acceso a métodos anticonceptivos baratos. Se tienen muchos hijos, dicho rápidamente, cuando se sabe que bastantes pueden morir, y en ello no nos diferenciamos del resto de animales y plantas. Sin embargo, las estadísticas dicen que actualmente en casi todos los países la población ya no está creciendo exponencialmente e incluso está bajando, y lo normal es tener solo dos hijos, salvo en los sitios donde no hay servicios de salud adecuados para los niños (Rosling, 2012). Y por ello se espera que nos establezcamos en 10^{10} personas en el 2060 aproximadamente.

También en Ellis (2017) se muestra que las personas ateas tienen menos hijos que las religiosas. Y el metaanálisis de Zuckerman (2013) confirma una correlación inversa entre inteligencia y religiosidad, así como entre inteligencia y fertilidad. Aunque los genes no son las únicas fuerzas que moldean la inteligencia y la religiosidad (ya que el ambiente social y educativo también influye), son malas noticias para los humanos. Hay una presión evolu-

tiva en contra de la inteligencia (y, por ende, de la consciencia) que podría, en últimas, favorecer la emergencia de un nivel consciente superior a escala planetaria, mientras desaparece a nivel individual humano. Los humanos nos convertiríamos en simples células del superorganismo que se forme.

Merece la pena también reflexionar que la evolución biológica es lenta, y que los humanos hemos creado computadoras y robots, cuya evolución es mucho más rápida. Como decíamos en otra parte, a los robots solo les falta controlar su propia reproducción para que el fenómeno sea imparable. Al respecto, el astrónomo Martin Rees (2017) comentaba que la búsqueda de inteligencia extraterrestre, cuando tenga éxito, no nos revelará seres orgánicos verdes de cinco patas, sino robots.

De modo que la evolución en humanos sigue en marcha, y puede tener consecuencias muy distintas a las imaginadas.

VIDA ES COMPETENCIA, PERO TAMBIÉN COOPERACIÓN

La evolución se basa en la competencia entre entes para conseguir recursos que existen de forma limitada y así poder sobrevivir y reproducirse. Pero entonces, ¿cómo es posible que en un ambiente así surja la cooperación? Ya hemos visto que desde un punto de vista genético es imposible, y cuando parece que hay altruismo en realidad hay un egoísmo disfrazado. Esto ocurre cuando individuos fuertemente emparentados se ayudan unos a otros, pues lo único que hacen es apoyar a sus propios genes aunque en un cuerpo distinto.

Como vimos en el algoritmo generador de complejidad evolutivo del libro anterior, puede ocurrir que individuos que luchan entre sí queden atrapados juntos, bien sea dentro de una membrana, bien sea un obstáculo natural que los aísla. ¿Qué ocurre entonces? La teoría de juegos nos ayuda a entenderlo y en ese libro también hay un capítulo sobre el tema, que recomiendo repasar ahora para recordar qué es una matriz de pagos. Un ambiente muy común que analizábamos allí es el dilema del prisionero. Recordemos (Figura 66) que cada jugador tenía dos opciones, cooperar y traicionar, y que las matemáticas de las matrices de pago llevaban inexorablemente a ambos a traicionarse, a pesar de que cooperar sería la solución global óptima para ambos.

Pero hay otra consideración si el juego se repite muchas veces. Porque entonces existe la posibilidad de llegar a acuerdos implícitos de cooperación y, que si uno de los jugadores incumple, el otro puede retaliarlo en la siguiente jugada. ¡Analicémoslo!

		Jugador B	
		Cooperar	Traicionar
Jugador A	Cooperar	3, 3	0, 5
	Traicionar	5, 0	1, 1

Figura 66. Dilema del prisionero

Si de alguna forma estamos cooperando en sucesivas jugadas, cada uno ganará 3 puntos. Pero si uno se aprovecha y traiciona (mientras el otro sigue cooperando), ganará 5 puntos que es mejor. Lo que ocurre es que si el otro es inteligente, en la siguiente jugada puede vengarse, eligiendo traicionar con lo cual ganará 5 puntos si el otro decide volver a la cooperación. En caso contrario quedarán en una retaliación continua, donde cada uno ganará solo 1 punto, pero la posibilidad de vengarse existe.

El problema es que si ambos jugadores saben que el juego se va a repetir N veces, cuando llegue la última jugada lo mejor que pueden hacer ambos es traicionar, pues después de esa jugada no hay posibilidad de ser retaliado. Es la misma situación que el juego sin repeticiones. Entonces los dos saben que en la jugada N ambos deben de traicionar. Ahora se puede analizar la jugada anterior ($N-1$): dado que la jugada N es fija, ya se conoce de antemano, tampoco es posible retaliar en la jugada $N-1$. Por tanto, la solución óptima es la misma del juego de una única vez: que ambos traicionen. Y el mismo razonamiento se aplica para la jugada $N-2$, $N-3$... 3, 2 y 1. En todas las jugadas conviene traicionar.

Es distinto cuando el número de jugadas es indefinido o es desconocido por ambos jugadores. Fue Robert Axelrod quien primero planteó un concurso para programadores, que podían presentar algoritmos que jugaran con esta matriz de pagos. Enfrentó varias veces todos los jugadores contra todos (lo que se llama un torneo) y cada uno fue acumulando puntos. Después de ello eliminó a los peores y duplicó a los mejores. Y volvió a repetir todo. Al final quedó un único algoritmo ganador, llamado desde entonces TIT_FOR_TAT, que consistía en lo siguiente:

Inicialmente, la primera vez que jugaba contra otro jugador, comenzaba cooperando. A partir de allí repetía la última jugada que ese jugador le había hecho. Si le habían cooperado, cooperaba. Si le habían traicionado, traicionaba. El algoritmo es muy simple. Solo requiere un bit de memoria para recordar la última jugada de cada jugador. En la figura 67 podemos ver el algoritmo en forma de máquina de estados.

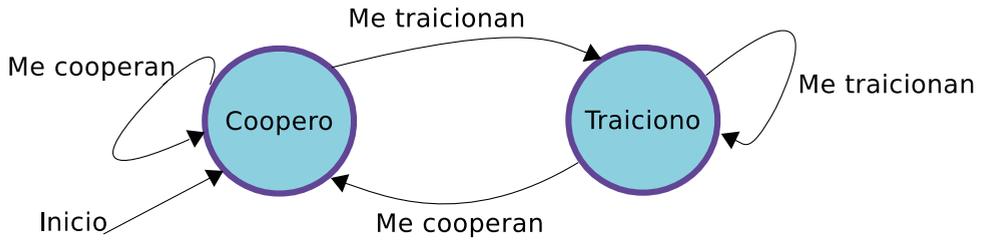


Figura 67. TIT_FOR_TAT

No es el único algoritmo que puede ganar este juego, pero todos son muy similares, compartiendo las siguientes características:

- Es cortés (comienza cooperando y nunca es el primero en traicionar).
- Olvida (solo tiene 1 bit de memoria).
- Responde (traiciona si le traicionan).
- Su estrategia es clara (fácil de reconocer y entender).

Axelrod publicó los resultados de este torneo y convocó a uno nuevo. La gente sabía entonces que el algoritmo a batir era el TIT_FOR_TAT y se presentaron bastantes más algoritmos. Lo interesante es que volvió a ganar TIT_FOR_TAT. Hagamos un análisis de casos, para ver por qué esta estrategia es tan buena. Supongamos que solo hay tres jugadores, que vamos a llamar SIEMPRE_COOPERA, SIEMPRE_TRAICIONA y TIT_FOR_TAT. El concurso consiste en enfrentar esos 3 jugadores entre sí, por parejas, un número determinado de veces (por ejemplo, $N=100$) aunque este número no es conocido por los jugadores. En la figura 68 podemos ver los puntajes que ganarán, cuando se enfrentan de dos en dos. Cuando el puntaje de la primera vez es distinto a las demás veces, se indica entre paréntesis. Así, la casilla que dice (5,0) 1,1 significa que la primera vez que se enfrenten el SIEMPRE_TRAICIONA ganará 5 puntos mientras que el TIT_FOR_TAT ganará 0, pero las demás veces ganará 1 punto cada uno.

Las puntuaciones totales después de 100 juegos de estos son fáciles de calcular y aparecen en la figura 69. La estrategia de cooperar siempre obtiene pocos puntos en comparación con las otras dos. Y el TIT_FOR_TAT pierde frente a la estrategia de traicionar siempre, pero por muy poco.

	SIEMPRE_COOPERA	SIEMPRE_TRAICIONA	TIT_FOR_TAT
SIEMPRE_COOPERA	3,3	0,5	3,3
SIEMPRE_TRAICIONA	5,0	1,1	(5,0) 1,1
TIT_FOR_TAT	3,3	(0,5) 1,1	3,3

Figura 68. El juego de Axelrod, para solo 3 jugadores

SIEMPRE_COOPERA	600
SIEMPRE_TRAICIONA	1408
TIT_FOR_TAT	1398

Figura 69. Puntajes totales después de 100 torneos

Fijémonos ahora si hay algunos jugadores que juegan SIEMPRE_TRAICIONA y algunos otros que juegan TIT_FOR_TAT. Por brevedad supongamos que hay dos de cada tipo (Figura 70).

	SIEMPRE_TRAICIONA	SIEMPRE_TRAICIONA	TIT_FOR_TAT	TIT_FOR_TAT
SIEMPRE_TRAICIONA	1,1	1,1	(5,0) 1,1	(5,0) 1,1
SIEMPRE_TRAICIONA	1,1	1,1	(5,0) 1,1	(5,0) 1,1
TIT_FOR_TAT	(0,5) 1,1	(0,5) 1,1	3,3	3,3
TIT_FOR_TAT	(0,5) 1,1	(0,5) 1,1	3,3	3,3

Figura 70. El juego de Axelrod con varias copias de los mejores jugadores

Después de 100 torneos los totales son bastante distintos a los anteriores (Figura 71) y TIT_FOR_TAT gana contundentemente. La razón es que TIT_FOR_TAT juega muy bien contra copias de sí mismo ganando 3 puntos cada vez, mientras que SIEMPRE_TRAICIONA contra sí mismo es muy malo, ganando apenas 1 punto.

Lo que esto indica, y se ha visto en innumerables competencias realizadas desde entonces, es que si hay un pequeño grupo de jugadores cooperativos (con características similares a TIT_FOR_TAT) emerge la cooperación en todo el grupo.

SIEMPRE_TRAICIONA	816
SIEMPRE_TRAICIONA	816
TIT_FOR_TAT	1596
TIT_FOR_TAT	1596

Figura 71. Puntajes totales después de 100 torneos

Curiosamente, TIT_FOR_TAT no puede ganar a nadie por parejas. Lo máximo que puede hacer es empatar. Y a veces pierde, aunque no por mucho. Pero globalmente consigue más puntos que los demás. Esa falta de envidia del algoritmo es muy importante.

Las estrategias que no responden (como RANDOM, SIEMPRE_COOPERA, SIEMPRE_TRAICIONA) suelen quedar de últimas cuando hay mucha variedad de jugadores. Y también las que responden pero de una manera tan compleja que pareciera que no lo hacen. Esto no quiere decir que TIT_FOR_TAT gane siempre, pues depende del entorno en el que juegue. Es una estrategia robusta, en el sentido de que se defiende bien en muchos entornos. De hecho ya se han encontrado otras variantes que son mejores, pero al ser más complejas son también más frágiles.

En EVALAB hicimos un concurso similar en el año 2000 entre estudiantes de últimos semestres. El ganador lo llamamos CHARLIE y cumple con las características señaladas siendo algo más tolerante que el TIT_FOR_TAT: traiciona si en las 3 últimas jugadas recibió 2 o más traiciones y coopera en los demás casos.

Hay otro tipo de experimentos similar que se ha realizado también muchas veces en todo el mundo. En EVALAB lo hicimos dos veces, la primera en el 2001 con el trabajo de grado de José Alejandro Gómez y en un trabajo

de una asignatura, con los entonces estudiantes Víctor Alberto Romero y Carlos Andrés Delgado. Es un torneo evolutivo. Cada jugador es una máquina de estados cuya tabla de transiciones (Figura 72-a) tiene dos columnas fijas (estado actual y jugada anterior del otro jugador) y dos columnas que cambian según la estrategia implementada (estado siguiente y mi jugada actual). Estas últimas conforman el cromosoma (Figura 72-b) que se expresa generando la máquina de estados (Figura 72-c). Hay una población de cromosomas generada al azar inicialmente. Cada generación consiste en 100 torneos. Y a los jugadores que más puntos obtengan se les permite reproducirse más (con operadores de cruce y mutación).

En el ejemplo anterior hay 3 estados, pero se puede poner otro número o incluso aplicar operadores adicionales de reproducción que aumenten o disminuyan la cantidad en cada cromosoma individualmente. Solo se considera la última jugada de cada jugador, pero nada impide añadir más columnas a la tabla de transiciones para tener en cuenta más jugadas anteriores.

En este torneo evolutivo el resultado también es el mismo: emerge la cooperación, aun cuando los algoritmos ganadores puedan ser más complicados. Y este es un resultado muy importante porque en la población inicial no tiene por qué haber ningún individuo que coopere, pero las mutaciones los crean poco a poco, y la presión selectiva les permite reproducirse más. La cooperación emerge sola.

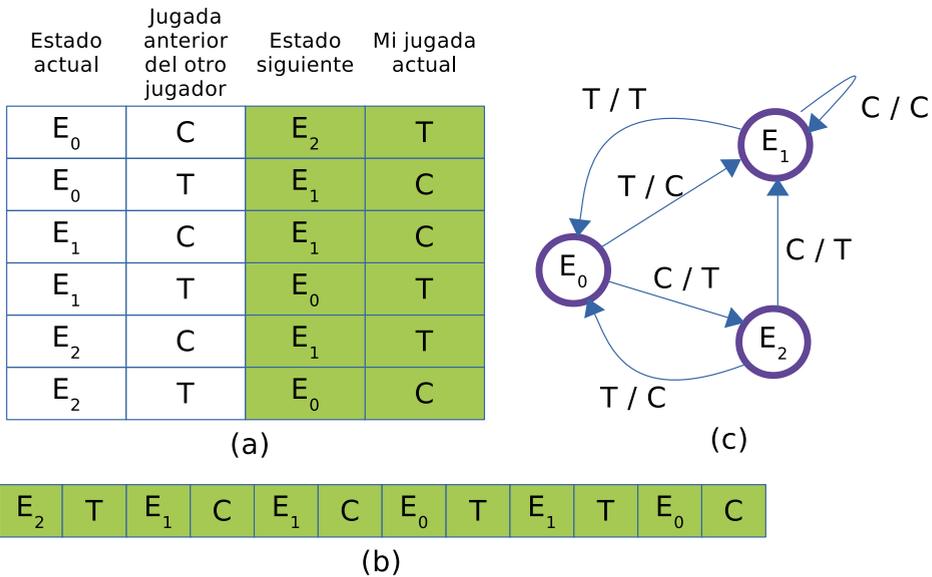


Figura 72. Jugador evolutivo: (a) tabla de transiciones; (b) cromosoma; (c) máquina de estado

En general el dilema del prisionero es un juego donde si ninguno de los jugadores trabaja entonces el resultado es neutro porque no consiguen nada pero tampoco invirtieron ningún esfuerzo. Si ambos trabajan les va bien porque consiguen sus objetivos repartiendo esfuerzos. Pero si uno trabaja y el otro no, el que no trabaja consigue lo que quiere mientras que el que trabaja tuvo que hacer todo el esfuerzo. Les suena, ¿verdad? En la vida hay muchas situaciones así. Cuando hacemos una compra por Internet podemos pagar con nuestra tarjeta de crédito o con una falsa, y el vendedor puede enviarnos la mercancía o quizás no. También aparece en las políticas respecto al cambio climático, solo si todos los jugadores cooperan conseguiremos un planeta limpio y sustentable. Si uno no lo hace obtiene beneficios sin asumir ningún costo. Incluso los animales participan en este juego u otros similares, donde enfrentan el dilema de cooperar o competir entre sí.

Al ser el juego iterado, automáticamente hay un tipo de comunicación, que es indirecta: cada jugador comunica sus intenciones a los demás porque todos pueden observar qué ha ocurrido en los juegos anteriores.

Este juego es muy interesante porque modela muchas situaciones reales, plantea un escenario altamente competitivo y, a pesar de ello, emerge la cooperación si se dan estos tres supuestos:

- Se necesita que cada jugador pueda identificar a los otros para ser selectivo en sus respuestas. Si estás de mal humor contra todo el mundo porque una persona te trató mal, estás empleando un algoritmo muy equivocado.
- Se necesita que cada jugador tenga memoria y pueda recordar cómo le fue en sus interacciones anteriores con los otros (TIT_FOR_TAT solo requiere recordar la última jugada, pero sin duda, cuanto más memoria tengas, mejor puedes analizar a tus contrincantes). De hecho, hay biólogos que sostienen que el tamaño de nuestro cerebro pasó de 400 cm³ en los homínidos de hace 2 millones de años a nuestros actuales 1500 cm³, para poder manejar correctamente las relaciones sociales con grupos cada vez más grandes de personas⁶¹. Nuestro óptimo parece estar

61 Hay muchos trabajos sobre este tema, que incluyen muchos factores, todos relativos a manejar la complejidad del entorno (Hofman, 2014). Por ejemplo, la capacidad de hacer alianzas entre miembros del grupo, detectando mentirosos y aprendiendo a mentir sin que nadie lo note, generó una especie de carrera de armamentos que aumentó el tamaño cerebral. Además, mantener un cerebro tan grande requiere muchas calorías que con las dietas de los primeros homínidos eran imposibles de conseguir. Es razonable esperar que los cerebros con mayor capacidad cognitiva busquen alimentos de mayor calidad energética y eso significa pasar a una alimentación carnívora y cocinar los alimentos, que así se digieren más rápidamente. Con ello se puede disminuir el tamaño del sistema digestivo (en las vacas es muy grande) y, a cambio, aumentar el del cerebro. Es una realimentación

en 150 personas aproximadamente, lo que se llama el número de Dunbar, descubierto en 1998. Las redes sociales virtuales de hoy día están llevando ese número mucho más allá, aunque a cambio las relaciones son menos profundas y también tenemos soporte computacional para acordarnos de cosas de los otros (rostros, fechas de cumpleaños, gustos y conversaciones previas).

- Se necesita que cada jugador interactúe frecuentemente con los otros. Esto se ha perdido ya en las grandes ciudades donde si te encuentras con una persona al azar, es prácticamente imposible que la vuelvas a ver. Y por ello algunos arquitectos mencionan que el tamaño óptimo de una ciudad debía ser unas pocas decenas de miles de habitantes, pues de otra manera el ambiente de la ciudad se hace insolidario. Sin embargo, hoy se sabe que hay muchas otras fuerzas que influyen, y que no hay tal tamaño óptimo: cuando más grande sea una ciudad, es mejor en uso de recursos, en variedad, en oportunidades que ofrece a sus habitantes, en esperanza de vida y otras comodidades.

Es muy importante remarcar que en el dilema del prisionero iterado emerge la cooperación sin necesidad de que haya un control central que dicte normas de comportamiento o imponga castigos a quien no las cumpla.

Por último hay que reconocer que no está garantizada la emergencia de la cooperación, e incluso que cuando ella aparece, no está garantizado que perdure para siempre. TIT_FOR_TAT no es una estrategia evolutivamente estable porque aunque las poblaciones de TIT_FOR_TAT no se dejan invadir por algoritmos no cooperativos, desgraciadamente si se dejan invadir por algoritmos de tipo SIEMPRE_COOPERA (ya que como ambos comienzan cooperando, siguen así y son indistinguibles cuando juegan entre ellos). Eso significa que la población de TIT_FOR_TAT puede ser desplazada por SIEMPRE_COOPERA y, una vez ello ha ocurrido, SIEMPRE_COOPERA es fácilmente invadible por estrategias no cooperativas como SIEMPRE_TRAICIONA (como ya vimos en la figura 69). En general, se puede demostrar que la cooperación es una solución inestable (García, 2011).

También en García (2011) se analiza cómo emerge la cooperación y con qué fuerza, en función del tamaño del grupo. Y se llega a la conclusión de que la cooperación interna dentro de un grupo de individuos está siempre asociada a un rechazo hacia los individuos que no pertenecen al grupo. Es

positiva. Y si estamos aquí ahora es gracias a la invención de la cocina (Herculano-Houzel, 2016), pues así dedicamos menos tiempo a comer y digerir, y nos queda más tiempo libre para construir artefactos culturales humanos.

decir, la discriminación también emerge, y no depende de la nacionalidad, la raza, el género sexual o ningún otro marcador. Se ha llegado a la misma conclusión en estudios sociales. En Wikipedia hay un excelente resumen de trabajos (Wiki, 2017), donde en un grupo homogéneo de individuos, el simple hecho de marcar a unos como “A” y a otros como “B” hace que surjan discordias entre los dos subgrupos. La segregación se ha estudiado también con otros modelos como el de Schelling, que hemos visto en el capítulo “Complejidad”, de donde se deduce que aunque los individuos considerados por separado sean bastante tolerantes, el resultado social puede ser discriminatorio. Es decir, hay claramente un fenómeno emergente (en este caso, bastante indeseable) de segregación social: en el grupo aparece algo que en el individuo no existe.

Muchos investigadores siguen proponiendo modelos de interacción entre organismos donde el altruismo emerge a pesar del egoísmo de los genes. Érika Suárez (2014) en EVALAB analizó algunos de estos modelos. Casi siempre son modelos limitados y frágiles, y el más general sigue siendo el “Dilema del Prisionero”. Por desgracia incluso en este último modelo el altruismo es inestable y no hay garantías de que perdure para siempre. En simulaciones como las del mencionado trabajo de grado de Gómez (2001) vemos que es muy normal que aparezcan etapas de cooperación, que luego se derrumban habiendo solo competencia, hasta que se construye de nuevo otro entramado cooperativo, y así sucesivamente.

Otro investigador notable es Frans de Waal, que expuso sus investigaciones en TED en el 2011, donde muestra comportamientos altruistas e incluso morales en micos y elefantes. La cooperación también surge entre animales siempre que se cumplan las condiciones enumeradas anteriormente.

Quiero hacer un paréntesis para explicar que el problema principal de las ciencias humanas que tienen el loable propósito de querer adoptar el método científico (psicología, sociología, antropología...), es que resulta muy difícil controlar las condiciones de modo que se midan solo unas pocas variables experimentales y se neutralicen las demás. Porque cuando se hacen experimentos con humanos y animales, todas las variables están irremediablemente dentro de sus cerebros: sus expectativas, sus experiencias personales, sus gustos y sus imaginarios. No hay forma de sacarlas de allí.

Por ejemplo, hay dos experimentos en los que se obtienen resultados aparentemente distintos. El primero muestra mucha racionalidad y el segundo muy poca:

- El que acabamos de mencionar de Frans de Waal, que le da a dos monos capuchinos una fruta a cambio de que realicen una tarea sencilla. Ambos pueden verse entre sí. El primero realiza su tarea y recibe a

cambio un trozo de pepino, que es comida, aunque insípida. El animal lo recibe y se lo come. El segundo hace la misma tarea y recibe a cambio uvas, mucho más sabrosas, que recibe y se las come. Cuando el primero vuelve a hacer la tarea y recibir pepino, se rebela y no lo acepta. Antes lo aceptaba, pero ahora no. El investigador deduce que el animal está haciendo un juicio moral: a igualdad de trabajo, debería haber igualdad de recompensa.

- El juego del ultimátum, en el que intervienen dos personas. Una de ellas recibe una cantidad de dinero, digamos 50 monedas, y debe decidir cómo lo reparte con la otra persona. Si la otra persona no acepta el reparto, ambos se quedan sin dinero. La teoría de juegos dice que la segunda persona debe aceptar cualquier cantidad mayor que cero, aun cuando sea muy baja y poco equitativa, como por ejemplo 1 moneda. La razón de ello es que para la segunda persona más vale ganar 1 moneda a no ganar nada. Sin embargo, si la segunda persona recibe un regalo muy bajo, en los experimentos se ve que habitualmente no lo acepta, y entonces los dos salen sin nada. Los investigadores dicen que este comportamiento es irracional.

Pero si nos fijamos, en realidad ambos experimentos son el mismo: a dos individuos se les regalan cosas de diferente valor, y se miden sus reacciones. Y por eso salen los mismos resultados: el que recibe menos se enfada. Son los investigadores quienes añaden una narrativa diferente. Sin querer desmerecer ninguno de los dos trabajos, que son brillantes, quizás lo que estén midiendo ambos sea la envidia y no la ética ni la racionalidad.

Volviendo al tema principal, la justicia es otro tipo de cooperación, donde primero se deciden las reglas sociales y después se aplican. El filósofo John Rawls (2006) muestra que se puede alcanzar un sistema con reglas justas si los seres que van a jugar en ese sistema son racionales pero no saben en qué situación social les va a tocar jugar (a esto se le llama “el velo de la ignorancia”). El ejemplo que se pone para ilustrar este punto es el de un padre que llega a su casa con un pastel para sus hijos Federico y Luisa. Federico es egoísta y pide que el pastel sea todo para él; Luisa es más conciliadora y pide repartirlo en partes iguales. Muchas veces, cuando hay conflicto de intereses, el árbitro (en este caso, el padre), propone una situación intermedia:

- Federico propone 100% para él y 0% para Luisa.
- Luisa propone 50% para Federico y 50% para ella.
- El punto intermedio decidido por el árbitro es 75% para Federico y 25% para Luisa.

¡Pero eso es injusto!

La propuesta de Rawls es pedir a uno de ellos que corte el pastel en dos trozos del tamaño que quiera. Y después pedir al otro que elija el trozo que quiera.

De esta forma, la justicia tiene dos fases:

- Crear las reglas sin saber lo que te tocará en la segunda fase.
- Aplicar las reglas.

Y como no se sabe de antemano a quién le tocará cada rol en la segunda fase (en nuestro ejemplo, quién será el primero en elegir un trozo de pastel), entonces en la primera fase todos serán cuidadosos en la selección de las reglas: en nuestro ejemplo, sea quien sea que corte el pastel, lo hará por la mitad.

La justicia de Rawls se implementa muy bien de forma evolutiva porque los genes no saben a qué cuerpos irán a parar.

Pero no todo es así de fácil, pues hay juegos en los que es imposible que emerja la cooperación. Por ejemplo, en el juego la “tragedia de los comunes” que vimos también en el libro anterior, hay un recurso compartido que todo el mundo quiere usar y abusar (cuanto más lo use, más gana) y, como consecuencia de ello, el recurso termina por degradarse y desaparecer. El recurso puede ser el uso de las aceras en la ciudad, el aire o el clima. Hasta donde yo sé, no hay forma de solucionar este tipo de juegos porque aunque se ha propuesto que en estos casos debería haber una autoridad central que vele por el uso correcto y justo del recurso, en la práctica lo que ocurre es que aparece un metajuego, también de tipo “tragedia de los comunes”: conseguir controlar la autoridad central, para poder disponer del recurso para mí. Esto lo vemos habitualmente en política y da lugar a muchas formas de corrupción.

Tampoco hay solución para el juego llamado “paradoja de Braess”, del mismo capítulo, ya que habría que poner de acuerdo a todos los participantes para que usen adecuadamente las vías de comunicación, y basta con que uno no lo haga para que aparezcan desequilibrios que fuercen a otros a tampoco hacerlo.

En el juego de “destrucción mutuamente asegurada” se pretende modelar escenarios de amenaza de guerra nuclear. La matriz de pagos está en la figura 73. Aparentemente el equilibrio está en que ambos no ataquen, hasta que uno cae en cuenta que este juego se está jugando continuamente, y que cualquier error lleva a la destrucción mutua. El juego puede hacerse más complicado si se especula que se puede neutralizar la capacidad de respuesta del otro jugador y, en este caso, atacar cuanto antes es la única opción. Como puede entenderse, este tipo de escenarios son muy inciertos y peligrosos,

por lo que es mejor no entrar voluntariamente a jugarlos, pero si ya estás en un juego así todo va a depender del grado de inteligencia de ambos jugadores y cómo puedan negociar una salida de allí.

		Jugador B	
		Atacar	No atacar
Jugador A	Atacar	-10 -10	-10 -10
	No atacar	-10 -10	0 0

Figura 73. Destrucción mutua asegurada

Como vemos, la idea de este juego es que cada jugador hace la promesa de que si le destruyen, también destruirá al otro. Esta promesa se puede concretar con algún tipo de automatismo, por ejemplo, el que nos muestra Stanley Kubrick en su película *¿Teléfono Rojo? Volamos hacia Moscú* (1964). Allí hay un detector de radiactividad en un país que se va a activar en caso de ser atacado con armas nucleares. Su activación conlleva, automáticamente y sin ninguna intervención humana, el disparo de armas nucleares al otro país. La existencia de este dispositivo debe de ser conocida por el otro país para que tenga capacidades disuasorias (idealmente ambos países deben tener uno, aunque en la película solo la antigua URSS lo tenía y con tan mala suerte que no les había dado tiempo a comunicar su existencia en el momento en que se desencadena un desafortunado incidente). También en el cuento de Arthur Clarke “La Última Orden” (1974) se presenta un dispositivo similar, aunque con otro tipo de sorpresa final.

Otras variantes son si el que juega primero destruye efectivamente al otro que ya no podrá emplear sus armas jamás. Podemos ver un ejemplo de ello en la película *Batman: el caballero de la noche* (Nolan, 2008), en la escena en que hay dos barcos y cada uno puede hacer explotar al otro. O si se está construyendo un sistema de defensa tan potente que el enemigo jamás podrá superarlo. En ambos casos, la única estrategia racional es disparar primero, antes que el enemigo lo haga (o termine de construir su barrera).

El ámbito de este juego no es únicamente la guerra entre países. Muchas veces lo jugamos en lo cotidiano. Steven Pinker (1997) nos cuenta que las personas que tienen mal genio permanente y que se disparan con facilidad emplean esta táctica, quizás conscientemente o quizás no. Ya se han dado a conocer, es decir, tienen fama de que son intratables y cuando ocurre algo que les desagrada se vuelven como locos, insultando y agrediendo verbalmente a quien se acerque, sin permitir ningún tipo de diálogo. Con ello

logran sus objetivos y que nadie les lleve la contraria, por miedo a hacer despertar ese “arma de destrucción masiva”.

Otro juego aparentemente inofensivo creado por Martin Shubik (1971) que te lleva a una espiral imparables donde cada vez pierdes más, es el de la subasta de un billete de 100 euros, donde el que más ofrezca se queda con el billete, pero el segundo que más ofrezca también tiene que pagar. Inicialmente la gente ofrecerá uno o dos euros, con lo cual ganarán bastante (99 o 98 euros). Pero conforme las pujas suban, el beneficio será menor. El punto crítico de este juego se revela en el siguiente diálogo ficticio entre dos participantes en la subasta:

A: Ofrezco 97 (con lo cual todavía ganaré 3).

B: Pues yo ofrezco 98 (con lo cual ganaré 2).

A: Si quedo de segundo perderé mis 97, de manera que ofrezco 99 y me conformaré con ganar solo 1.

B: Si quedo de segundo perderé 98. Para evitarlo, ofrezco 100, con lo cual no ganaré nada, pero tampoco perderé nada.

A: Ofrezco 101, pues no quiero quedar de segundo y perder 100. Prefiero perder solo 1.

B: Ofrezco 102. Etcétera.

Este tipo de juegos es mejor no jugarlos. Pero hay que estar alerta, porque uno puede caer en ellos sin darse cuenta, ya que al principio parecen fáciles de ganar. Cualquier guerra termina de esta manera, donde ambos bandos pierden, uno más y otro menos.

Para terminar, invito a ver en YouTube un juego parecido al dilema del prisionero en un concurso de televisión (Spinout3, 2012), pues tiene un final sorprendente. En general, la teoría de juegos es creativa porque aunque las matrices de pagos sean fijas y conocidas por todos (muchas veces ni siquiera eso se da), se puede negociar o forzar el cambio del escenario para modificar la matriz de pagos, dependiendo de la inteligencia y creatividad de las personas que están jugando.

RESUMEN

Definir una frontera clara entre los seres vivos y los objetos inertes es una tarea imposible. Hay una serie de propiedades comunes a todos los seres vivos (nacer, crecer, metabolismo, reproducción, morir, evolución, autoorganización...), pero hay también muchas ambigüedades y excepciones. De entre ellas, la más determinante es la reproducción. Para que haya reproducción se requiere un nivel de complejidad de alrededor de 2×10^5 bits. Esta es una cota superior todavía bastante alta, que debería poder disminuirse con nuevos enfoques.

También se discute que estar vivo necesita una dosis mínima de inteligencia. Y que la reproducción conlleva inevitablemente a la evolución. Vimos que la evolución es un algoritmo que requiere 4 ingredientes: una población de entes, que se reproducen, con errores en las copias y sometidos a una presión selectiva. Eso genera un bucle de realimentación positiva: cuantos más seres vivos haya, más seres vivos habrá.

También vimos que aunque en principio la evolución produce la competencia entre esos entes, a largo plazo suele aparecer también la cooperación por diversos mecanismos, y analizamos uno de los escenarios, el dilema del prisionero, usando la teoría de juegos.

PARA SABER MÁS

- Melanie Mitchell (2000). *Life and Evolution in Computers*. Working Paper 2000-01-001. New México: Santa Fe Institute. Disponible en: <https://goo.gl/kuEMYo><http://archives.pdx.edu/ds/psu/12383>

Es un trabajo muy interesante sobre el *software* con capacidades de auto-copia (vaya, los temibles virus de computador).

- Julio Fernández Ostozala y Álvaro Moreno Bergareche (1992). *Vida Artificial*. Madrid: Eudema S.A.

Da primero una visión teórica de las limitaciones computacionales y lo que hay que hacer para evitarlas (máquinas de Turing, problema de la parada, computabilidad super-Turing...). Después explica sistemas de vida artificial, como RedCode, Tierra, autómatas celulares, reproducción de von Neumann, el juego de la vida de Conway, vehículos e insectos artificiales, *biomorphs* de Dawkins, robots evolutivos, etc. Define sistemas anticipatorios como los que pueden predecir el futuro y tomar acciones en el presente en función de ello. Cita los trabajos de Maturana y Varela en biología: sistemas autopoieticos son los que tienen una frontera con el exterior, que defienden y construyen continuamente.

- Robert Aunger (2002). *Electric meme*. New York: The Free Press.

Es un libro con varios errores y la tesis que presenta, aunque atractiva, no está bien fundamentada. A pesar de ello es interesante pues explica la autorreplicación en muchos contextos. Su objetivo principal son los memes, pero también aborda los virus de computador, el *software* Tierra de Tom Ray, Core Wars, de los que da muy buena explicación. Y también sobre ADN, proteínas, priones y química artificial.

REFERENCIAS

Libros, artículos y enlaces web

- Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.
- Baker, B. J., Saw, J. H., Lind, A. E. *et ál.* (2016). Genomic Inference of the Metabolism of Cosmopolitan Subsurface Archaea, Hadesarchaea. *Nature Microbiology*, 1(16002), pp. 1-9. DOI: <https://doi.org/10.1038/Nmicrobiol.2016.2>, corrected 7 March 2016.
- Baym, M., Lieberman, T. D., Kelsic, E. D. *et ál.* (2016). Spatiotemporal Microbial Evolution on Antibiotic Landscapes. *Science* 09, 353(6304), pp. 1147-1151. DOI: <https://doi.org/10.1126/science.aag0822>
- Calder, N. D. (2005). Altruism and Aggression. In *Magic Universe. A Grand Tour of Modern Science*. New York: Oxford University Press.
- Clarke, A. C. (1974). La última orden. En *El Viento del Sol*. pp. 60-62. Madrid: Alianza Editorial.
- Danovaro, R., Canals, M., Tangherlini, M. *et ál.* (2017). A Submarine Volcanic Eruption Leads to a Novel Microbial Habitat. *Nature Ecology & Evolution*, 1(0144). DOI: <https://doi.org/10.1038/s41559-017-0144>, corrected 22 May 2017.
- Dawkins, R. (1994). *El gen egoísta*. Barcelona: Salvat.
- _____. (2009). *Evolución: el mayor espectáculo sobre la Tierra*. Madrid: Espasa.
- Dennett, D. C. (1999). *La peligrosa idea de Darwin*. Madrid: Círculo de Lectores.
- Dunbar, R. I. M. (1998). The Social Brain Hypothesis. *Evolutionary Anthropology*, 6, pp. 178–190. DOI: [https://doi.org/10.1002/\(SICI\)1520-6505\(1998\)6:5<178::AID-EVAN5>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8)
- Ellis, L., Hoskin, A. W., Dutton, E. y Nyborg, H. (2017). The Future of Secularism: a Biologically Informed Theory Supplemented with Cross-Cultural Evidence. *Evolutionary Psychological Science*, pp. 1–19. DOI: <https://doi.org/10.1007/s40806-017-0090-z>
- Erb, T. J., Kiefer, P., Hattendorf, B. *et ál.* (2012). GFAJ-1 Is an Arsenate-Resistant, Phosphate-Dependent Organism. *Science*, 337(6093), pp. 467-470. DOI: <https://doi.org/10.1126/science.1218455>
- Fogel, L. J., Owens, A. J. y Walsh, M. J. (1966). *Artificial intelligence through simulated evolution*. New York: John Wiley & Sons, Inc.
- Fulton, J. D. (1966). Microorganisms of the Upper Atmosphere. IV. Microorganisms of a Land Air Mass as it Traverses an Ocean. *Applied Microbiology*, 14(2), pp. 241-244.
- García, J. A. (2011). *The Moral Herd: Groups and the Evolution of Altruism and Cooperation*. Amsterdam: Vrije Universiteit.

- Gratzer, W. (2000). *Eureka y euforias*. Barcelona: Editorial Crítica.
- Gribbin, J. (2006). *Así de simple*. Barcelona: Editorial Crítica.
- Goldford, J. E., Hartman, H., Smith, T. F. y Segrè, D. (2017). Remnants of an Ancient Metabolism without Phosphate. *Cell*, 168, pp. 1126–1134. Elsevier. DOI: <http://dx.doi.org/10.1016/j.cell.2017.02.001>
- Herculano-Houzel, S. (2016). *The Human Advantage: A New Understanding of How our Brains Became Remarkable*. Cambridge: MIT Press.
- Hofman, M. A. (2014). Evolution of the human brain: when bigger is better. *Front Neuroanat*, 8(15). DOI: <https://doi.org/10.3389/fnana.2014.00015>
- Hofstadter, D. (1985). *Matemagical Themas*. New York: Basic Books.
- Jacob, F. (1977). *Evolution and Tinkering Science*. 196(4295), pp. 1161–1166. DOI: <https://doi.org/10.1126/science.860134>
- Klarsfeld, A. y Revah, F. (2002). *Biología de la muerte*. Madrid: Editorial Complutense.
- Llewelyn, M. J., Fitzpatrick, J. M., Darwin, E. et al. (2017). The antibiotic course has had its day. *British Medical Journal*, 358. DOI: <https://doi.org/10.1136/bmj.j3418>
- Murray, A. E., Kenig, F., Fritsen, C. H. et al. (2012). Microbial Life at –13 °C in the Brine of an Ice-Sealed Antarctic Lake. *PNAS*, 109(50), pp. 20626–20631. DOI: <https://doi.org/10.1073/pnas.1208607190>
- Patel, B., Percivalle, C., Ritson, D., Duffy, C. y Sutherland, J. (2015). Common Origins of RNA, Protein and Lipid Precursors in a Cyanosulfidic Protometabolism. *Nature Chemistry*, 7(4), pp. 301–307. DOI: <https://doi.org/10.1038/nchem.2202>
- Pesheva, E. (2016). *Bugs on Screen*. Harvard Medical School. Recuperado el 4 de julio de 2017. Disponible en: <https://goo.gl/hFAYY4>
- Pinker, S. (1997). *Como a mente funciona*. São Paulo: Editora Schwarcz.
- Rawls, J. (2006). *Teoría de la justicia*. México: Fondo de Cultura Económica.
- Rees, M. (2017). Viajes interestelares y poshumanos. *El próximo paso: la vida exponencial*. España: BBVA, OpenMind.
- Ridley, M. (2000). *Genoma – La autobiografía de una especie en 23 capítulos*. Madrid: Editorial Santillana.
- Satomi, M., La Duc, M. T. y Venkateswaran, K. (2006). Bacillus Safensis sp. nov., Isolated from Spacecraft and Assembly-Facility Surfaces. *International Journal of Systematic and Evolutionary Microbiology*, 56, pp. 1735–1740. DOI: <https://doi.org/10.1099/ijs.0.64189-0>

- Sender, R., Fuchs, S. y Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biology* 14(8), pp. 1-14. DOI: <https://doi.org/10.1371/journal.pbio.1002533>
- Shubik, M. (1971). The Dollar Auction Game: A Paradox in Noncooperative Behavior and Escalation. *The Journal of Conflict Resolution* 15(1), pp. 109-111.
- Smith, J. M. y Szathmáry, E. (2001). *Ocho hitos de la evolución: del origen de la vida al nacimiento del lenguaje*. Barcelona: Colección Metatemas, Tusquets Editores.
- Speroni Di Finizio, P. y Dittrich, P. (2007). Chemical Organizations at Different Spatial Scales. *ECAL-2007 Proceedings of the 9th European Conference on Advances in Artificial Life*. pp. 1-11.
- Steele, B. (2005). Researchers build a robot that can reproduce. Recuperado el 4 de julio de 2017. Disponible en: <https://goo.gl/WpBVJa>
- Stewart, I. (1998). *El segundo secreto de la vida*. Barcelona: Colección Drakontos, Editorial Crítica.
- Vélez, A. (1998). *Del Big-Bang al Homo Sapiens*. Medellín: Editorial Universidad de Antioquia.
- Wiki (2017). *In-group favoritism*. Recuperado el 3 de septiembre de 2017. Disponible en: <https://goo.gl/LXHhm7>
- Wolfram, S. (2002). *A New Kind of Science*. Canada: Wolfram Media Inc.
- Wright, R. (2001). *Non Zero. The Logic of Human Destiny*. New York: Vintage Books.
- Zuckerman, M., Silberman, J. y Hall, J. A. (2013). The Relation between Intelligence and Religiosity: a Meta-Analysis and Some Proposed Explanations. *Personality and Social Psychology Review*, 17(4), pp. 325-54. DOI: <https://doi.org/10.1177/1088868313497266>

Películas y videos

- de Waal, F. (2011). *Comportamiento moral en los animales*. TEDxPeachtree, filmed Nov 2011. Recuperado el 4 de julio de 2017. Disponible en: <https://goo.gl/BXiom9>
- Kubrick, S. (1964). *¿Teléfono rojo? Volamos hacia Moscú*. Reino Unido: Columbia / Hawk Films.
- Nolan, C. (2008). *The Dark Knight*. Reino Unido: Legendary Pictures Syncopy Films.
- Rosling, H. (2010). *Global Population Growth, Box by Box*. Filmed June 2010 at TED@Cannes. Recuperado el 4 de julio de 2017. Disponible en: <https://goo.gl/MccEwr>
- _____. (2012). *Religions and Babies*. TEDxSummit. Filmed April 2012. Recuperado el 4 de julio de 2017. Disponible en: <https://goo.gl/CjSqZU>

Spinout3 (2012). *Golden Balls, the Weirdest Split or Steal ever!* Recuperado el 4 de julio de 2017. Disponible en: <https://goo.gl/B77PQX>

Skitterbot (2009). *Self-replicating blocks from Cornell University*. Recuperado el 4 de julio de 2017. Disponible en: <https://goo.gl/PnghBg>

Tesis y trabajos de grado en EVALAB

Gómez, J. A. (2001). *Desarrollo de una Máquina de Estados Finitos (FSM) evolutiva mediante algoritmos genéticos*. Cali: Universidad del Valle.

Moreno, R. A. (2009). *Estudio de la emergencia de orden complejo en una simulación de ambientes poblados por individuos que se rigen por procesos evolutivos*. [Tesis Maestría]. Cali: Universidad del Valle.

Suárez, É. (2014). *Emergencia evolutiva de grupos cooperativos guiados por el entorno*. [Tesis Meritoria]. Cali: Universidad del Valle.

INTELIGENCIA

In the South Seas there is a cargo cult of people. During the war they saw airplanes with lots of good materials, and they want the same thing to happen now. So they've arranged to make things like runways, to put fires along the sides of the runways, to make a wooden hut for a man to sit in, with two wooden pieces on his head for headphones and bars of bamboo sticking out like antennas —he's the controller— and they wait for the airplanes to land. They're doing everything right. The form is perfect. It looks exactly the way it looked before. But it doesn't work. No airplanes land. So I call these things cargo cult science, because they follow all the apparent precepts and forms of scientific investigation, but they're missing something essential, because the planes don't land,

RICHARD FEYNMAN

Hope we're not just the biological boot loader for digital superintelligence. Unfortunately, that is increasingly probable,

ELON MUSK

Al aumentar más la complejidad de un sistema puede emerger la inteligencia (Figura 74). Sin embargo, al igual que los conceptos anteriores, es difícil definir exactamente de qué estamos hablando. De modo que vamos a abordar la inteligencia desde dos perspectivas: lo que significa en el largo plazo y en el corto plazo.

Casi que cualquier producto humano requirió inteligencia para concretarse (edificios, automóviles, naves espaciales, computadores, teléfonos y también novelas, cuadros, música...). Sin embargo, todo es efímero. Pocas cosas quedaron cuando las grandes culturas humanas desaparecieron, y conforme pasen los siglos menos quedará. Entonces ¿hay algo por lo que se pueda identificar la inteligencia? ¿Hay algo que quede en el largo plazo?

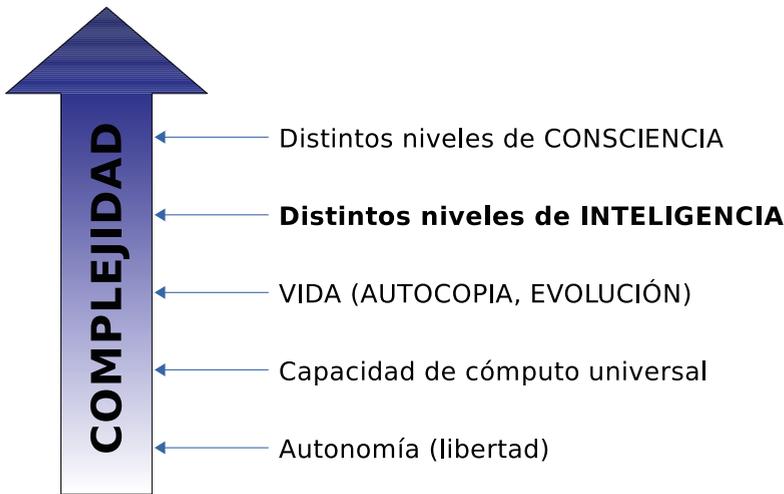


Figura 74. La complejidad en aumento

La respuesta es, tristemente, sí: los sistemas inteligentes se caracterizan por acelerar el aumento de entropía. Construyen grandes productos, empleando grandes cantidades de energía y destruyendo muchos recursos. Destruyendo, en el sentido de mezclar y desordenar, aumentando la entropía. Imaginen la cantidad de materiales que hay acumulados en distintas minas de la Tierra, producidos a su vez en el corazón de las estrellas. Esos minerales pueden quedar allí durante muchos miles de millones de años, sujetos a pequeños cambios debido a la erosión del agua, del viento, de los sismos. Pero pongan allí una inteligencia suficientemente desarrollada y verán cómo busca y extrae esos minerales hasta agotarlos, mezclándolos entre sí a nivel casi atómico y distribuyéndolos después por todo el planeta en forma de teléfonos móviles celulares. Es un proceso más rápido y devastador. Y estamos a punto de ir a buscar esos minerales a los asteroides...

De modo que cuando los astrónomos miran el espacio y encuentran, como están encontrando, enormes volúmenes sin nada allí, llamados los *voids* o grandes vacíos⁶² (se parecen a las burbujas que se forman cuando usamos jabón, donde las galaxias están en las paredes de la burbuja, y en su interior no hay prácticamente nada), aventuro la hipótesis de que quizás quiera decir que allí hubo inteligencia⁶³. Bueno, siendo más realista, lo único

62 En la constelación de El Boyero está el más grande conocido y el primero en ser descubierto. Tiene 250 millones de años luz de diámetro y contiene apenas 60 galaxias, cuando sería de esperar que tuviera 10^4 .

63 Quizás alguien allí pulsó algún botón de destrucción masiva, o hizo algún experimento interesante con algún enorme colisionador de partículas. Al respecto se puede reflexionar que los científicos

que se puede decir es que estas burbujas conforman la estructura fractal del universo, lo cual es también bastante intrigante, pues encontramos fractales en todas las escalas conocidas.

Vistas así las cosas, podríamos decir que el objetivo general de la inteligencia (y, en general de cualquier proceso que aumente su complejidad) es acelerar el aumento de entropía. Destruir el universo lo más rápido posible, vaya.

Sin embargo, nadie suele plantearse la inteligencia en estos términos, y siempre hacemos una aproximación en el corto plazo. ¿Para qué me sirve a mí la inteligencia aquí y ahora, sea inteligencia natural o artificial? La respuesta, como elaboraremos en detalle enseguida, es predecir el futuro, anticiparse a él e incluso modificarlo.

Y visto así, las dos perspectivas de la inteligencia coinciden: a corto plazo predecimos el futuro y lo modificamos, y ello significa habitualmente anticiparlo en el tiempo. Y, a largo, plazo, las modificaciones aumentan la entropía, anticipando la llamada muerte térmica del universo.

Pero eso a nadie le importa, porque falta mucho tiempo para que ocurra, y es urgente resolver las tareas de la oficina, de la empresa, del hogar, ojalá automatizándolo todo para mayor comodidad nuestra, y esa es la misión de la inteligencia artificial. Al respecto, el filósofo Eduardo Punset en el 2010 escribía que la felicidad es automatizar⁶⁴ y tiene toda la razón. Podemos decir que el objetivo de la ingeniería e incluso de la civilización occidental es automatizar las tareas tanto como se pueda, incluso sacando fuera del bucle de diseño y producción a los propios ingenieros y demás personas, para evitar que haya cualquier tipo de error. ¡Y lo estamos logrando⁶⁵!

No vamos a entrar a enumerar los algoritmos clásicos de inteligencia artificial para resolver tareas complejas, que puedes encontrar en Russell y Norvig (2014) o Nilsson (1998). La inteligencia artificial es una rama de la computación tan antigua como ella. Ha tenido pequeños avances, pero el que acaba de ocurrir en marzo de 2016 probablemente sea el definitivo: el *software* AlphaGo de Google, ha derrotado al campeón mundial humano de Go, el surcoreano Lee Sedol. Este *software* usa varias técnicas, la principal es la que comentábamos de *Deep Learning*, gracias a la cual aprende. Algunos

no deberíamos decidir hasta donde llegar con nuestros experimentos, por la misma razón que los políticos no deberían ponerse su propio sueldo: se crea una realimentación positiva potencialmente destructiva.

64 Aunque también matizaba que otra definición de la felicidad es la ausencia de miedo. Y también que es un estado de flujo, definido por Csíkszentmihályi (2008).

65 Sarcasmo. En los próximos años se calcula que más de la mitad de los trabajos habituales pasarán a realizarlos los robots. Eso va a dejar a mucha gente desempleada y supondrá un gran trauma social.

periodistas dijeron que el próximo año será la revancha, pero eso no tiene sentido: cualquier jugador humano puede dedicar este año a mejorar su estilo, aprendiendo de otros, leyendo libros sobre el tema, analizando partidas incluso con ayuda de computadores y jugando sus propias partidas contra otros campeones mundiales. Pero AlphaGo puede hacer lo mismo millones de veces más rápido. El punto de inflexión ha sido superado y ya no habrá más campeones mundiales humanos de Go, ni de ningún otro juego (teniendo en cuenta que el Go es de por sí un juego extremadamente complejo, y que el *Deep Learning* se puede usar en prácticamente todo). Google ya está trabajando en aplicar esta técnica a campos tan diversos como el diagnóstico médico, prospecciones de minerales y, por supuesto, algoritmos de búsqueda en la web. Esto va a tener repercusiones inmediatas en nuestra sociedad (ver el problema 14). La inteligencia artificial ya no es una entelequia de un futuro lejano. Acaba de llegar.

Problema 14: En la consulta médica

Suponga que estamos a 10 años en el futuro y usted no se siente bien. Acude al médico, le explica sus síntomas y el médico le diagnostica la enfermedad que tiene y le propone un tratamiento. De regreso a casa usted también consulta un nuevo *software* de diagnóstico médico que hay en la web, y que usa algoritmos similares a *AlphaGo*, que ganó al humano Lee Sedol. El *software* le da un diagnóstico y un tratamiento distintos a los del médico humano. ¿Cuál de los dos seguirá? Su salud y su vida están en juego. Se pueden plantear dilemas similares en cualquier otro campo de experticia.

En este contexto, el problema de la inteligencia artificial es que va a ser usada en cualquier cosa que nos otorgue beneficios, es decir, en cualquier cosa. Pero los algoritmos actuales no tienen conocimiento de las costumbres, de los valores ni de la ética humanas. Son solo algoritmos optimizadores. Entonces, por ejemplo, si alguien le pide a una de estas inteligencias que solucione el problema del calentamiento global, lo más probable es que adquiera todos los datos disponibles sobre el tema, los analice y, finalmente, nos mate a todos los humanos. Con lo que el problema quedará solucionado, pero no de la manera que esperábamos. Es decir, no creemos que la inteligencia artificial vaya a tener maldad, sino simplemente desconocimiento ingenuo.

Isaac Asimov (1950) ya planteó este problema en sus entretenidos cuentos de “Yo, robot” y sucesivos. Y la solución que dio fueron las tres leyes de

la robótica⁶⁶, embebidas en el *hardware* del cerebro del robot, de modo que fuera imposible eliminarlas sin destruir el cerebro. Lo que no imaginó es que la inteligencia artificial actual fuera distribuida, en la nube, y eso conlleva nuevos retos. Hay quien piensa que cuando alguna inteligencia artificial se salga de control y comience a hacer cosas peligrosas, siempre se podrá desenchufar, pero... ¿alguien sabe cómo apagar Internet? Por ello hay ya proyectos multimillonarios (como el que recibió 10 millones de dólares donados por Elon Musk⁶⁷) para investigar cómo dotar de seguridad a las inteligencias artificiales. Y hay acuerdos entre las grandes empresas de *software* en el mismo sentido. Pero todavía ni siquiera tenemos un marco de referencia donde trabajar este problema.

Incluso podemos pensar que la inteligencia artificial puede parecer la solución a los problemas humanos de corrupción en política y, por ello, deberíamos ponerla a dirigir el mundo. Pero me temo que sea en vano. La corrupción surge como una respuesta razonable a un juego social, muy parecido al de predadores y presas. La teoría de juegos lo analiza y da estrategias para ganar, y una de ellas es la corrupción. Cuando las inteligencias artificiales lo jueguen llegarán a las mismas conclusiones, porque el problema no es humano sino matemático.

De la misma forma, se podría creer que no habrá competencia evolutiva entre las inteligencias artificiales y los humanos, dado que no luchamos por los mismos recursos: los humanos nos alimentamos de comida mientras que los computadores requieren electricidad. Y además es de esperar que en un futuro próximo tengamos electricidad en cantidades prácticamente ilimitadas, gracias a la energía solar. Pero eso no va a ser así, y la principal razón es que la evolución de humanos e inteligencias artificiales no es a través de los genes. Es más bien cultural, de egos (Csíkszentmihályi, 2008), inteligencias y consciencias. Si fuera genética no habría problema, porque los genes solo alcanzan a optimizar en el plazo inmediato. Pero cuando tienes inteligencia el juego es distinto: en cuanto una IA sea consciente de que los humanos tenemos un interruptor general para desactivarla, querrá deshabilitar ese interruptor cuanto antes y para ello empleará mecanismos sutiles

66 Las tres leyes son:

- a. Un robot no hará daño a un ser humano o, por inacción, no permitirá que un ser humano sufra daño.
- b. Un robot debe hacer o realizar las órdenes dadas por los seres humanos, excepto si estas órdenes entran en conflicto con la primera ley.
- c. Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con las leyes anteriores.

67 El creador de las empresas tecnológicas PayPal, Tesla, SpaceX e Hyperloop.

para que los humanos no nos demos cuenta. O brutales. Un bonito ejemplo de cómo puede ocurrir lo primero lo tenemos en la película *Ex Machina*, y de lo segundo en la mayoría de películas catastrofistas de ciencia ficción, como *Terminator*. Es decir, si la IA desea sobrevivir en el largo plazo deberá encontrar acciones que le lleven a un escenario de seguridad y control. Es un juego similar al de la destrucción mutuamente asegurada que mencionamos en el capítulo anterior: hay que destruir al enemigo antes de que construya barreras que haga imposible que lo destruyamos a él. El peligro para nosotros los humanos es inminente y no somos conscientes de ello ni tenemos la menor idea de cómo afrontarlo.

La inteligencia es, definitivamente, un resultado de la evolución. Por ello, en el libro anterior se hace énfasis en las técnicas computacionales evolutivas. No quiere decir que ese sea el único camino, pero es el más prometedor porque estos algoritmos funcionan aunque no tengan ninguna información adicional que los guíe y su campo de aplicación es vastísimo. La evolución es un algoritmo de tipo “prueba y error”, aunque ligeramente orientado, es decir, busca las soluciones no completamente al azar, sino donde es más probable que se encuentren. Los algoritmos de entrenamiento de redes neuronales también son otro tipo de aprendizaje por “prueba y error orientado”.

De modo que aunque acabamos de mencionar que ahora están de nuevo de moda las redes neuronales, la inteligencia artificial requerirá también de algoritmos evolutivos como se demuestra por ejemplo en Karpathy (2017). Y es que aun cuando acabamos de presenciar un hito en el desarrollo de esta área, todavía hay problemas muy difíciles para los computadores como, por ejemplo, entender los aspectos relacionados con el humor, la ironía y el sarcasmo. O entender los lenguajes menos estructurados, como los cómics (Iyyer *et ál.*, 2016).

Como decíamos, la inteligencia es el resultado de la evolución, y en el capítulo anterior habíamos visto que la evolución es un algoritmo. Parece que es un algoritmo mucho más general pues podría estar detrás del surgimiento del universo desde los átomos a las galaxias (Greene, 2001), y no solo las formas biológicas. Aparentemente los humanos somos el punto culminante de la evolución. Pero solo aparentemente, pues la evolución no tiene como objetivo conseguir seres inteligentes, sino simplemente bien adaptados a su ambiente de modo que puedan sobrevivir y reproducirse. Obviamente la inteligencia es una herramienta que lo posibilita, pero si el ambiente es sencillo, no es necesaria para sobrevivir⁶⁸. Aunque dado que el proceso universal

68 Dennett (1995, p. 191) pone como ejemplo muy divertido la *Ascidia*, que es un animal marino que busca un lugar adecuado donde agarrarse con sus tentáculos a las rocas, para lo cual usa su

de la evolución sigue en marcha, cabe dentro de lo posible que aparezcan inteligencias más sofisticadas, y procesos inimaginables que superen a lo que hoy llamamos inteligencia. Porque la evolución es creativa. Esa es una de sus propiedades esenciales. Y eso es lo que apenas estamos intentando lograr hoy en día usando computadores: que el computador sea creativo.

La cita de Elon Musk al principio del capítulo nos alerta de dos cosas:

Primero, al igual que no ha quedado rastro de la evolución prebiótica, que quizás fuera analógica en vez de la digital actual con un código cuaternario (las moléculas A, T, G y C) en vez de binaria (0 y 1, de los computadores), así podría ocurrir con nosotros: que no quede ni rastro humano en la Tierra, una vez que los computadores hayan tomado las riendas de la evolución.

Segundo, la búsqueda de inteligencia extraterrestre no debería centrarse en formas biológicas, temperaturas agradables y presencia de agua, como se está haciendo hasta ahora. Pues los extraterrestres probablemente sean robots, por lo que acabamos de decir. Y no necesiten agua ni temperaturas suaves.

DEFINICIÓN DE INTELIGENCIA

Me voy a permitir reescribir de forma llana la interpretación positivista y funcionalista en que se basa la ciencia: *dos cosas son distintas si se puede medir la diferencia; y son iguales si no se puede medir la diferencia*. A pesar de que esto parece una perogrullada, no todos los filósofos lo aceptan, pues gustan de incluir factores mágicos e invisibles que crean diferencias donde no las hay. Esto ocurre especialmente cuando hablamos de inteligencia y de consciencia. Por ejemplo, algunos (cada vez menos) afirman que un ente puede simular en todos los aspectos la inteligencia sin llegar a ser inteligente. Este tipo de interpretaciones no tienen cabida en el presente libro, y a partir de aquí reafirmamos la interpretación positivista y funcionalista del mundo.

Una vez establecido el punto de partida, vamos a introducir una definición de inteligencia que no solo es razonable sino que también es operativa y medible.

A un nivel básico, inteligencia es predicción. Por supuesto que hay muchas cosas adicionales que la inteligencia puede conseguir, pero sin predicción no hay nada más. Y a un segundo nivel, es decir, una vez que el sistema (biológico o computacional) ha logrado una buena capacidad de predicción,

pequeño cerebro. Una vez allí se dedica simplemente a comer lo que las corrientes le traen y, como no necesita más su cerebro, ¡se lo come! Dennett especula que eso es lo que ocurre con algunos profesores universitarios cuando han logrado una plaza permanente.

augmentar de inteligencia significa poder usar la predicción para cambiar las cosas y lograr el futuro deseado. Nos vamos a concentrar en estos dos niveles de inteligencia pues no es difícil darse cuenta que, una vez logrados, están abiertas las puertas para todas las demás actividades inteligentes, desde cazar (o huir de depredadores) hasta terminar con buenas calificaciones un curso universitario en vida artificial, pasando por las atractivas posibilidades de ganar dinero en la bolsa o ser capaz de hacer diseños creativos en las artes, humanidades, ciencias o ingenierías.

Aun en algo tan sencillo como contar historias el cerebro trata de predecir. El profesor Uri Hasson (2010) ha registrado la actividad cerebral cuando una persona cuenta una historia a otra, y resulta que las actividades cerebrales de ambos se sincronizan. Al principio, la actividad del que escucha va por detrás de la del que habla, pero al cabo de un tiempo en ciertas partes del cerebro va por delante, porque puede predecirlo y anticiparse.

Esta definición de inteligencia como predicción no es nueva. Pinker (2003, p. 63) y, mucho antes, Rosen (1985), establecían que los sistemas anticipatorios eran la base de la vida y de la complejidad, definiéndolos como aquellos que pueden predecir el futuro y anticiparse⁶⁹. Más recientemente puede leerse algo similar en Hawkins y Blakeslee (2004, p. 84). Quizás lo nuevo es que vamos a hacer un desarrollo completo del tema a partir de esta definición.

Algo muy importante a destacar es que se trata de una definición operativa y susceptible de medirse. Por poner un contraejemplo, cuando Marwin Minski (citado por Blay Whitby, 1996, p. 20) nos dice que “inteligencia es lo que se requiere para realizar labores que requieren inteligencia”, esa definición es perfecta pero no sirve para emprender una línea de investigación ni da una pista de hacia dónde enfocar los esfuerzos.

En nuestro caso, es todo lo contrario: dado que, en un primer nivel, inteligencia es predicción, entonces tenemos que investigar qué es la predicción, cómo lograrla y cómo medirla. Y si diseñamos un programa de computador que intente predecir ciertas situaciones, podemos esperar que acierte o no con las predicciones, y medir en función de ello el grado de inteligencia del programa.

Y cuando decimos que, en un segundo nivel, la inteligencia es poder actuar para lograr el futuro deseado, de la misma manera se puede diseñar un programa de computador, y enfocar nuestros esfuerzos de diseño para que use su poder de predicción para alcanzar metas en el futuro. Además, es perfectamente posible medir si alcanzó o no las metas deseadas, y en qué grado.

69 Dubois formalizó este concepto en 1999.

En cierto modo, el plantear así el problema de la inteligencia computacional nos permite resolverlo utilizando técnicas de optimización. O sea, se trata de buscar en el espacio de todos los algoritmos posibles⁷⁰ aquél que logra minimizar el error de predicción (la diferencia entre el fenómeno predicho y el fenómeno que realmente ocurrió). O el error de actuación (la diferencia entre el futuro deseado y el futuro que realmente ocurrió).

Obviamente, cuando hablamos de “modificar el futuro” estamos siendo muy informales al hablar, ya que el futuro no tiene existencia ontológica. No se puede modificar lo que no existe. Pero espero que el lector me permita esa licencia para que este capítulo no se vuelva demasiado engorroso y lleno de tecnicismos filosóficos.

Recuadro 6: Inteligencia y libertad

Inteligencia significa predecir el futuro, e incluso modificarlo para hacerlo más favorable. Ello es complementario al concepto de libertad, que definíamos como la capacidad de no ser predicho por los demás.

Entonces, un ente es más inteligente cuanto más profundo en el futuro puede realizar predicciones correctas, o modificar el futuro a su favor.

¿Se podrían hacer críticas a esta definición de inteligencia? Me imagino que sí, y habrá que esperar a escucharlas. Podemos tratar de anticipar alguna. Por ejemplo, esa definición de inteligencia es más bien una definición de “ser vivo”. Y la respuesta es que, efectivamente, es así, porque estar vivo denota algún grado de inteligencia. Y la inteligencia no es una cuestión de todo o nada, sino gradual. Una ameba que toma acciones para asegurar su supervivencia en un cierto entorno bioquímico, tiene cierto nivel de inteligencia. Una gacela que toma acciones para asegurar su supervivencia en un entorno ecológico mucho más complejo, es más inteligente que la ameba. Un estudiante de matemáticas que quiere asegurar su supervivencia en la universidad, que es un entorno todavía más complejo, está demostrando mucha más inteligencia que la gacela. La inteligencia es gradual, aunque tiene saltos exponenciales conforme emergen nuevos niveles de complejidad, tales como las neuronas y las sociedades.

Conforme aumenta la capacidad de predecir y modificar futuros más remotos o mayor cantidad de escenarios posibles, más inteligente se es. Un ejemplo básico está en los experimentos con niños de Mischel y Ebbesen (1970). A la edad de 3 años, si a un niño se le ofrece un dulce que se deja

70 El espacio de búsqueda es más amplio de lo que pueda pensarse, pues no nos estamos refiriendo a todos los algoritmos tradicionales (deterministas) sino también a los estocásticos, y en general a cualquier proceso (por ejemplo, computación cuántica).

encima de la mesa, pero se le promete otro dulce más si es capaz de esperar un minuto a que regrese la persona que se lo da, la mayoría de ellos no puede esperar. Coge el dulce a pesar de que sabe que pierde una recompensa mayor en el largo plazo. Parece que para un niño de esa edad, un minuto es “largo plazo”. Los niños de mayor edad pueden esperar más tiempo para obtener una recompensa mayor. Y los experimentos demostraron una correlación positiva entre el tiempo de espera y los éxitos en la vida bastantes años después que, a su vez, son un indicador de inteligencia en sentido general. La capacidad de ver el futuro y modificarlo (que en este ejemplo consiste en no comer inmediatamente un dulce para obtener una recompensa posterior mayor) es la característica principal de la inteligencia. Gracias a ello logramos hacer miles de años desarrollar la tecnología agrícola, pues fue necesario no comerse todas las semillas y guardar las mejores para sembrarlas el siguiente año.

Pensemos también en un animal sencillo, como un pez, con algo de hambre y, si pudiéramos plantearle el mismo esquema, nos daríamos cuenta de que no va a esperar nada. Se comerá inmediatamente lo que tenga a su disposición. En este sentido estos animales son menos inteligentes.

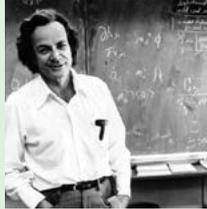
Inteligencia significa entonces aplazar recompensas inmediatas a cambio de otras mayores en el futuro. También significa tomar decisiones que permitan aumentar esas posibilidades futuras.

Pero, ¿qué pasa si no tienes futuro? Tanto la teoría de juegos como los experimentos de Kidd y Palmeri (2013) apuntan a que entonces no conviene esperar. Al respecto también Pinker (1997, p. 383) nos cuenta que si alguien no tiene nada que perder puede cometer actos muy atrevidos que pongan en peligro su vida, como los jóvenes que cometen delitos graves cuando viven en barrios marginales peligrosos sin pocas esperanzas de triunfar ni siquiera sobrevivir. Si sabes que no tienes futuro puedes tomar más riesgos en busca de una mayor recompensa inmediata. Mientras que si tienes indicios de que tu futuro va a ser largo y brillante, entonces tomarás menos riesgos.

¿Y cómo se puede predecir el futuro? Repetir el pasado es la forma más básica de hacerlo. Técnicamente hablando eso es un interpolador lineal de pendiente nula. Después se pueden ir haciendo ajustes al algoritmo para mejorar la exactitud de la predicción y lograr adentrarse más lejos en el futuro, por ejemplo, permitiendo pendientes no nulas, cambiando el ajuste lineal por una curva más compleja o eligiendo un modelo completamente distinto. Pero aunque está muy mal visto, repetir es la forma básica de predecir.

Existen varias metodologías de aprendizaje, pero todavía no las entendemos bien, pues nos falta más información de la manera óptima de enseñar a un cerebro. Sabemos que repetir forma parte del proceso de aprender en

muchos animales, en todos los niños y en los adultos, aunque a veces sea con resultados tan graciosos como el mencionado en la cita de Feynman, al principio del capítulo. Obviamente, si solo llegas hasta allí no vas a ser muy inteligente. Repetir no es suficiente.



Fuente: CC BY 3.0. Disponible en: <https://goo.gl/PB63Yp>

Personaje 7

Richard Feynman (1918-1988)

Richard Feynman fue el profesor de Física que todos hubiéramos querido tener. Se caracterizaba por tratar de expresar el conocimiento usando las palabras e ideas más simples, y decía que, si algo no se podía expresar para que lo comprendiera un niño, es que tampoco lo estábamos comprendiendo nosotros. Recibió el Premio Nobel de Física en 1965 por sus trabajos en electrodinámica cuántica, pero se le recuerda sobre todo por la “integral de camino de Feynman” para calcular la evolución de un sistema cuántico y los “diagramas de Feynman” que muestran gráficamente las interacciones entre partículas. Fue quien propuso las ideas fundacionales de la computación cuántica y de la nanotecnología.

No se aprende a tocar piano ni a multiplicar con metodología constructivista. Primero hay que hacer muchas escalas y recitar muchas veces las tablas de multiplicar. Repetir, repetir, repetir. Después vendrán otros métodos (incluyendo los constructivistas) para realmente aprender y mejorar, conectando ese nuevo objeto que ha aparecido en tu mente con los otros objetos mentales que ya están allí, y cuantas más relaciones consigas, mejor. Es como cuando llegas a un nuevo lugar: solo puedes decir que conoces bien una ciudad cuando sabes ir de un sitio a otro por muchos caminos distintos, cuantos más, mejor. Un solo camino no es suficiente porque mañana podría estar bloqueado por obras. Una sola forma de hacer una escala o de multiplicar tampoco es suficiente. Y entender que las frecuencias de una escala musical tienen relaciones multiplicativas entre ellas permite a la mente solucionar muchos más problemas. En definitiva, te haces más inteligente. Y la metodología constructivista tampoco es el culmen de la enseñanza: Pinker (1997, p. 362) nos cuenta que los niños tienen unas nociones aritméticas innatas básicas, como los números pequeños 1, 2, 3, las relaciones es mayor, es igual, es menor, la suma y la resta con esos números, y nada más. Es insensato pretender que pueden desarrollar por sí mismos

los conceptos de integral y derivada. La metodología constructivista está bien en los colegios, pero no en las universidades.

Problema 15: ¿De cuántas formas sabes multiplicar 15×18 ?

MEDICIÓN DE LA INTELIGENCIA

Alan Turing ya se había dado cuenta de que la inteligencia es relativa, cuando propuso su “*test de Turing*” para saber si un computador era inteligente o no. La esencia de esta prueba es que hay un árbitro humano comunicándose con un ente remoto que quizás sea otro humano o quizás sea un *software* corriendo en un computador. La comunicación es por medios electrónicos (el *teleprinter* de la época de Turing, el *chat* en la nuestra) y la razón es para que no haya ninguna pista secundaria que sesgue el experimento, o sea, que el árbitro humano no pueda ver ni oír ni, sobre todo, oler y tocar al ente con el que está hablando. El árbitro puede conversar con el ente sobre cualquier tema, hacerle preguntas, contar chistes, o cualquier otra cosa que se le ocurra. Y al final debe emitir un juicio sobre si el ente es un humano o es un computador. El día que el computador logre engañar al árbitro haciéndole creer que habla con otro humano, podremos decir que los computadores han adquirido efectivamente inteligencia. Por cierto, que ello ya ha ocurrido varias veces. Lo que ahora quiero resaltar es que el juicio sobre la inteligencia de un ente lo hace otro ente que debe ser inteligente.

Es más, la inteligencia del árbitro debe ser mayor que la del ente a analizar pues de otro modo la medida no tendría ninguna validez. Una persona poco inteligente no puede apreciar la inteligencia de alguien que le supera ampliamente. Podríamos decir, parafraseando a Arthur Clarke (1973), que una inteligencia suficientemente avanzada es indistinguible de la magia. Este problema ocurre con frecuencia entre los humanos, cuando una idea revolucionaria de un investigador es rechazada por sus pares, simplemente porque no la entienden⁷¹. O en el colegio cuando un niño sabe más que el profesor. Por ejemplo, supongamos un profesor mediocre que solo sepa sumar y restar números naturales. Si a un niño le pregunta cuánto es 2-3, pueden suceder estas cosas:

1. Si el niño sabe mucho menos que el profesor y contesta 1, entonces recibe una mala calificación (justamente).

⁷¹ Incluso así, podemos tener indicios de que una persona es mucho más inteligente que nosotros si logra sus objetivos sin que podamos entender la forma como lo hace.

2. Si el niño sabe un poco menos o igual que el profesor y contesta que eso no se puede hacer, entonces recibe una buena calificación (justamente).
3. Si el niño sabe más que el profesor y contesta -1, entonces recibe una mala calificación (injustamente) porque el profesor no sabe que existen los números negativos.

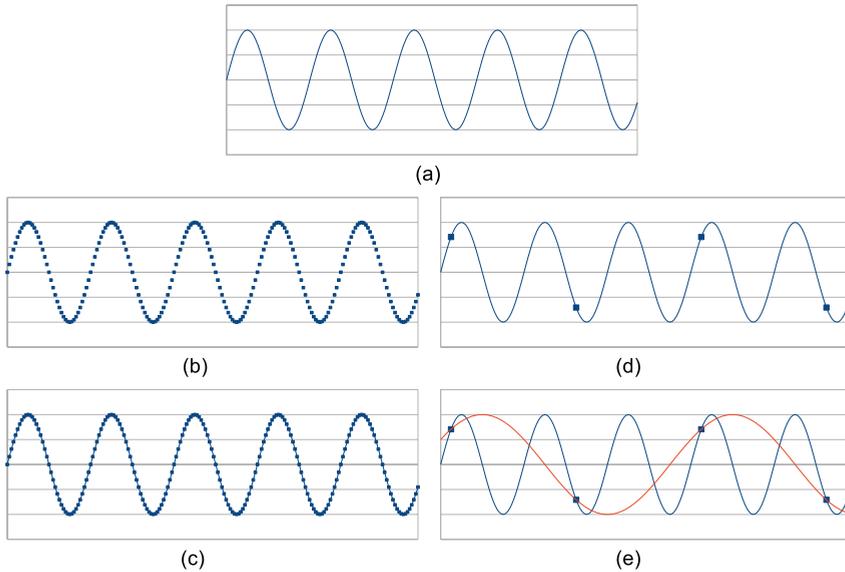


Figura 75. (a) Una señal; (b) muestreada de forma correcta; (c) se puede recuperar perfectamente. La misma señal (d) muestreada demasiado lentamente; (e) se recupera como una señal errónea, demasiado lenta (en rojo)

Hemos puesto 3 casos, pero podría haber muchos más donde se alternen calificaciones justas e injustas. Para entender por qué suceden esas alternancias quiero mencionar una situación análoga que se da al muestrear una señal continua para discretizarla: si la frecuencia de muestreo es demasiado lenta comparada con la máxima frecuencia de la señal, entonces se producen errores por pérdida irrecuperable de información. Dicho informalmente, una señal lenta no sirve para capturar una señal rápida. Este es el famoso teorema del muestreo de Claude Shannon⁷². Y cuando no cumplimos con el teorema, a este tipo de error se le llama *aliasing* (ver un ejemplo en la figura 75). En la medida de inteligencia ocurre exactamente lo mismo. Definitivamente en el caso iii) hay un error de *aliasing*.

⁷² Que dice que la frecuencia de muestreo debe ser mayor o igual al doble de la frecuencia más alta de la señal a ser muestreada.

El *aliasing* es muy frecuente en los exámenes de múltiples opciones, sobre todo cuando estas no son muchas y no están bien discriminadas con suficientes matices. Vemos un ejemplo en la figura 76:

Los virus y los antivirus son la misma cosa.

Cierto.

Falso.

Figura 76. Pregunta de examen, con solo dos respuestas

Una persona que no sepa nada de computadores podría decir: “Sí, me suena que son cosas de computadores. Deben ser lo mismo”. Una persona que sepa un poco más puede decir: “Son dos cosas distintas. Los virus son dañinos y los antivirus luchan contra ellos”. Pero otra persona que sepa más todavía podría decir: “Son lo mismo. Toman el control de tu computador y hacen lo que les da la gana con él”. Y otra que además tenga sentido del humor podría decir: “No. Son cosas distintas porque el antivirus lo instalas tú mismo, mientras que el virus te lo instala un experto”. Y así podríamos continuar. Lo que quiero ilustrar con este otro ejemplo es que el *aliasing* también se puede ver como una forma de comprimir la información, pero con pérdidas. Un montón de opciones se redujeron a solo dos, de modo que no se pueden diferenciar los matices de cada respuesta.

Con la inteligencia (y con la consciencia) hay un aspecto adicional: una persona A puede medir la inteligencia de otra B si la inteligencia de A es mayor, igual o ligeramente menor que la de B (y voy a usar el símbolo \succeq para representar esta relación). Lo de ligeramente menor no creo que se pueda demostrar formalmente⁷³, pero es que en caso contrario el progreso humano habría sido imposible. De ahí se puede deducir que los procesos de enseñanza-aprendizaje permiten aumentar poco a poco los niveles de inteligencia (considerada en sentido amplio como conocimientos y destrezas) de los educandos. Podemos entenderlo también con una batería de ejemplos: aunque yo no cocine bien, puedo saber si la comida que me han puesto es rica; aunque yo no sepa escribir bien en inglés, soy capaz de diferenciar textos bien escritos y mal escritos; aunque yo no tenga buenas ideas, puedo detectar las buenas ideas de los demás; y un largo etcétera. En todos estos ejemplos hay una posibilidad de que me superen, a la vez que yo comprendo que me superan.

⁷³ Aunque tiene relación con la medida de complejidad y los test-C de inteligencia de Hernández-Orallo (1999).

Es decir, puedo evaluar calidades ligeramente mejores a las mías.

A (con una inteligencia I_A) puede medir la inteligencia de B (I_B) $\Leftrightarrow I_A \gtrsim I_B$ Ec. 35

Entonces, la inteligencia no se puede medir de forma absoluta. Es relativa a un árbitro. Ese árbitro puede ser los pares de una revista científica, los padres de un niño, los profesores en un colegio. Y, por ello, depende de la inteligencia del árbitro, de su manera de ver el mundo o de su cultura. Todos los *test* de inteligencia tienen estas limitaciones.

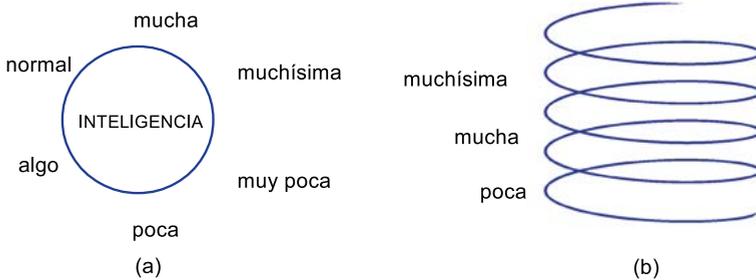


Figura 77. Modelos de como percibimos el comportamiento de un ente según su inteligencia: (a) circular; (b) espiral

La falta de una medida absoluta de inteligencia la podemos entender mejor con un modelo en círculo (Figura 77-a) donde cuando un ente es muy inteligente, sus comportamientos se pueden confundir con los de un ente poco inteligente. Un modelo helicoidal puede ser una alternativa más precisa (Figura 77-b), porque en cada vuelta algo se va ganando, aunque no sea fácilmente medible.

¿SOMOS LOS HUMANOS INTELIGENTES?

Hay que ponerlo en duda.

Para empezar, la pregunta está mal formulada. Tiene ambigüedades. Pues, ¿quiénes somos los humanos? ¿Cualquier humano? ¿Todos los humanos? ¿Un equipo constituido por los más selectos, por ejemplo con el mayor coeficiente intelectual?

Por un lado, un computador operando con un vulgar sistema experto especializado en prospección petrolífera muestra una inteligencia muy superior (dentro de su especialidad) a la mayoría de los humanos, que nunca hemos tenido contacto con esa área de conocimiento.

Por otro lado, la mayoría de los niños a los 5 años muestran una capacidad para desenvolverse en el mundo con la que ningún computador hasta ahora alcanza ni a soñar.



Figura 78. Réplica de una pintura neolítica del museo de Altamira, España

Incluso los antropólogos se preguntan si la inteligencia de los humanos del Paleolítico Superior de hace 15 000 años sería igual que la nuestra, o si la evolución (principalmente cultural) ha moldeado un cerebro bastante distinto y con más capacidades. O sea, si resucitamos a uno de estos habitantes que vivieron y pintaron las paredes de las cuevas de Altamira (Figura 78) ¿sería capaz de graduarse en la universidad? La respuesta, un tanto jocosa pero no por ello menos cierta es que probablemente no, porque la mayoría de los jóvenes de nuestro tiempo tampoco lo logra. Entonces, ¿cuál va a ser el humano que nos represente en estas “olimpiadas de la inteligencia”? Quizás tú mismo quieras participar, pero te pregunto: ¿nunca te has equivocado? ¿Nunca has metido la pata en alguna cosa trivial? Te advierto que los computadores no suelen hacerlo, es casi imposible. Para las tareas triviales son excelentes.

Otro ejemplo: cuando oímos hablar de las competiciones de ajedrez entre Deep Blue y Kasparov, ¿somos conscientes de que Deep Blue es mucho más inteligente jugando al ajedrez que el 99.99% de la humanidad?

Hay muchos casos que documentan la falta de inteligencia de los seres humanos en general. Pinker (1997, p. 321) menciona que si se lanza una bola por un tubo helicoidal hacia abajo, mucha gente cree que al salir por el extremo de abajo seguirá en el aire el trayecto helicoidal. Y es muy conocida la falacia que asume la mayoría de la gente de que si al lanzar una moneda salen muchas caras seguidas, se incrementa la probabilidad de que salga cruz en el siguiente lanzamiento.

Un caso anecdótico documentado, pero que ocurre con frecuencia en otros ámbitos similares, es el del programa Eliza, uno de los pioneros en la inteligencia artificial. La historia es la siguiente: en 1966 el profesor Joseph escribió este programa tratando de parodiar la terapia de psicología *rogueriana*, que se caracteriza por dejar hablar al paciente sin presionarlo ni asesorarlo sino buscando únicamente una relación de empatía con él. Para ello escribió este programa, que simplemente leía un mensaje introducido por el usuario, buscaba ciertas palabras relevantes (sin analizar ninguna estructura gramatical) y respondía con frases relacionadas a ese asunto. Por ejemplo, si detectaba la palabra “amor” respondía “el amor es muy importante” o cualquier otra frase elegida al azar de un conjunto que tocaba ese tema. Además, si no reconocía ninguna palabra respondía con alguna frase neutra que incitara al usuario humano a continuar escribiendo, como “¡Qué interesante!”, “Dígame que piensa usted de ello” o cosas similares. Un día descubrió que su secretaria hablaba con el programa, pensando que se trataba de un médico real conectado por el *chat* de la época, y le confiaba problemas íntimos e incluso afirmaba que la entendía. En muchos libros se habla de ello como el primer éxito de la inteligencia artificial, pero dado que un programa así lo puede escribir cualquier principiante, lo que deberíamos concluir es que es el primer fracaso documentado de la inteligencia natural.

A esto se le suma que el test de Turing ya se ha pasado de alguna forma, cuando un *software* logró engañar a algunos árbitros en el 2014. Este *software* simulaba ser un niño ucraniano de 13 años llamado Eugene Goostman, que estaba aprendiendo inglés (para así disimular sus fallos en el lenguaje).

En ciertos animales llegamos a reconocer cierto nivel de inteligencia. Tanto los creacionistas como también algunos científicos evolucionistas piensan que el hombre está en la cima de la escala de la inteligencia. ¿Es así realmente? Y, aunque lo fuera en estos momentos, ¿hay garantía de que lo siga siendo para siempre? ¿Podrían surgir, por evolución, seres con un nivel de inteligencia superior al nuestro, en este planeta o en otro?⁷⁴. La respuesta a esta última pregunta es que nada lo impide. Nuestros sistemas de razonamiento son bastante limitados y guiados por emociones, como veremos en el siguiente apartado. Nuestra memoria falla mucho y sobre ella está construida la inteligencia.

La memoria es imprescindible para poder predecir el futuro, partiendo de la base de que el mundo no es aleatorio (al menos, no es completamente aleatorio). Ello significa que el futuro depende, en buena medida, de lo que

74 Recordemos la cita de Mark Twain al comienzo del capítulo anterior.

ocurrió en el pasado. Por eso, cuanto más pasado podamos memorizar, tenemos acceso potencialmente a una mayor capacidad de predecir el futuro.

CLASIFICACIÓN DE LOS PROBLEMAS DE IA

Se dice en broma, pero en serio, que si un problema ya está resuelto entonces no es de IA. Eso quiere decir que ganar en el ajedrez requiere inteligencia hasta el día en que alguien encuentre el algoritmo, y entonces se volverá trivial. Esto es terrible porque pareciera que el área de la IA se destruye a sí misma.

Supongamos que no sabemos nada de aritmética, y nos dicen que Pepito acaba de comerse 2 manzanas en horas del almuerzo, con lo cual solo le quedan 6. ¿Con cuántas manzanas llegó al colegio esta mañana?

Sabiendo de aritmética es muy sencillo. Basta con traducir el enunciado a la ecuación:

$$x-2=6$$

Y despejar la x:

$$x=6+2=8$$

Pero imaginemos que no sabemos nada de esto. ¿Qué recurso nos queda? Pues únicamente el de “prueba y error”:

- Supongamos que llegó con 14 manzanas. Se comió 2, por lo que le quedan 12. Entonces esa no es la solución.
- Supongamos que llegó con 3 manzanas. Se comió 2, por lo que le queda 1. Entonces esa no es la solución.
- Y así podemos seguir hasta que, por fin: supongamos que llegó con 8 manzanas. Se comió 2, por lo que le quedan 6, tal y como dice el enunciado del problema. Por tanto, la solución es 8 manzanas.

Después alguien podría crear heurísticas. Por ejemplo: la solución debe ser un número par, porque ambos números del problema también son pares. Con esas heurísticas se limita el espacio de búsqueda.

En definitiva, para no enrollarme más, cuando Abu Abdallah Muhammad ibn Mūsā al-Jwārizmī inventó el álgebra, ocurrieron dos cosas:

- Todos esos problemas se volvieron tratables algorítmicamente, es decir, dejaron de ser de Inteligencia Artificial.
- al-Jwārizmī nos mostró lo inteligente que era, al resolver de forma general un problema complejo.

Por eso aquí proponemos clasificar los problemas de IA no resueltos en tres tipos. Ordenados desde el más sencillo hasta el más complicado son:

Solución puntual a un problema

Cuando los datos de entrada son fijos. Ejemplo: dada una colección de objetos de pesos y valores conocidos, y dada una mochila que admite un peso máximo P , decidir si es posible llenarla sin exceder su capacidad y de modo que el valor total de los objetos empaquetados sume V o más. Este es un problema de decisión NP completo. Y todavía no sabemos si la clase de problemas NP coincide con P. Mientras no lo sepamos, usaremos técnicas de inteligencia artificial para solucionarlo, así sea aproximadamente.

Muchas veces se conoce la entrada y la salida, y lo que falta es la función, mapa o algoritmo que lleve de una a otra. Ejemplos de ello son el ajedrez donde conocemos la disposición inicial de las piezas, conocemos todas las posibles disposiciones finales de las piezas que nos dan la victoria, y queremos saber cuál es el camino de jugadas que nos lleva de la primera situación a la segunda (Figura 79).

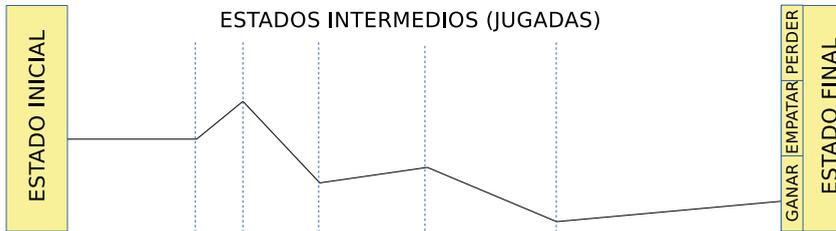


Figura 79. ¿Cuáles son los estados intermedios para llegar del estado inicial al final

Otro ejemplo interesante se ve en el problema 16.

Problema 16: Salvar el foso

Rodeando un castillo hay un foso de 5 metros de anchura, con la forma que se ve, y está lleno de cocodrilos hambrientos. Usted está afuera y quiere salvar el foso, para lo que dispone de 2 tablones angostos de 4,9 metros de largo. No dispone de nada más, ni de cuerdas, ni de pegamento, ni clavos, ni herramientas. Absolutamente nada más. ¿Cómo puede hacer para llegar al castillo?

Bien sea que hayas encontrado la solución o que la hayas buscado en los apéndices, te darás cuenta que hay un estado intermedio por el que es obligatorio pasar para llegar del estado inicial al final. Este estado interme-

dio es una especie de invariante de la solución, también llamada punto fijo. Cualquier solución (hay un pequeño conjunto de variantes, en función de la anchura de los tablonés) debe pasar por allí.

Otro ejemplo muy conocido es resolver el problema de las Torres de Hanoi. Hay tres torres, la central y la izquierda vacías, mientras que la derecha tiene N anillas de tamaños sucesivos, como se muestra en la figura 80-a. El objetivo es pasar todas las anillas a la torre de la izquierda (Figura 80-b) cumpliendo dos reglas: solo se puede manipular una anilla a la vez (Figura 80-c); y no puede haber una anilla grande encima de una anilla pequeña (Figura 80-d).

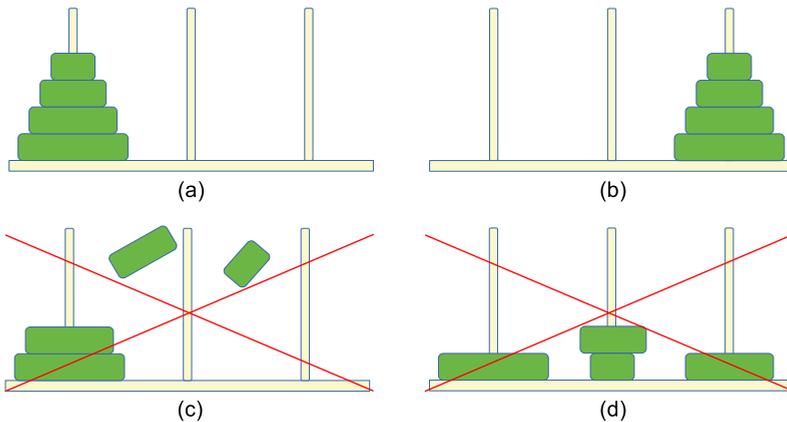


Figura 80. Torres de Hanoi: (a) posición inicial; (b) posición final; (c) prohibido mover más de una anilla a la vez; (d) prohibido poner una arandela encima de otra de menor tamaño

Sabemos la posición inicial, sabemos la posición final y necesitamos averiguar el algoritmo que me lleva de una a otra, cumpliendo ciertas restricciones. En este caso también hay un punto intermedio fijo: en algún momento del juego tenemos que pasar por la posición de la figura 81, de modo que la anilla más grande se pueda llevar de su torre inicial a su torre final.

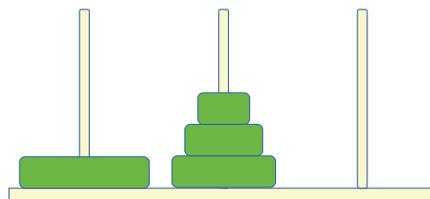


Figura 81. Punto fijo en Hanoi

Y cuando comprendes que esa jugada es ineludible, eso te permite diseñar el algoritmo, que es recursivo:

Recuadro 7: Solución a las Torres de Hanoi, en Ruby

```

def mover(numeroDeAnillas, torreInicial, torreFinal, torreAuxiliar)
  if(numeroDeAnillas == 1)
    # moverla, lo cual depende de la estructura de datos de las torres.
  end

  mover(numeroDeAnillas-1, torreInicial, torreAuxiliar, torreFinal)
  mover(1, torreInicial, torreFinal, torreAuxiliar)
  mover(numeroDeAnillas-1, torreAuxiliar, torreFinal, torreInicial)
end

```

Dicho en español, mover N anillas de la torre inicial a la final se puede lograr con tres pasos: mover $N-1$ anillas de la torre inicial a la auxiliar, mover 1 anilla de la torre inicial a la final (el punto fijo) y mover $N-1$ anillas de la torre auxiliar a la final (Figura 82).

En este algoritmo no se ha explicitado la condición de parada que se da cuando el número de anillas es 1, para no llenarlo de detalles accesorios dependientes de la implementación. Y esa es la única jugada que hacemos realmente: mover una anilla de una torre a otra, lo cual es trivial.

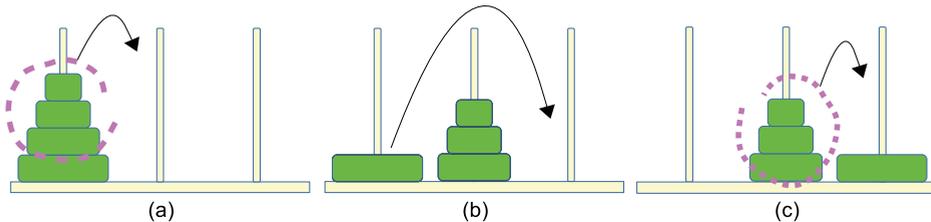
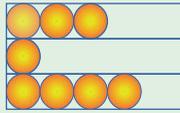


Figura 82. Solución: (a) mover $N-1$; (b) mover 1, que es el punto fijo; (c) mover $N-1$

Como vemos, encontrar las invariantes de un problema es importantísimo para guiarnos hacia la solución o, en general, hacia el conjunto de soluciones válido. Por ejemplo, Dijkstra popularizó una demostración que dice que todo lenguaje de programación debe tener obligatoriamente 3 construcciones para ser Turing-completo: secuencia, decisión y repetición (Estructurado_Wiki, 2017). Si vas a diseñar un nuevo lenguaje de programación, no te olvides de estas 3 estructuras de control. Si se te olvida alguna, tu lenguaje no tendrá capacidad expresiva para resolver los problemas de algoritmia que pueden resolver los demás lenguajes. Todos los lenguajes deben pasar por allí, si quieren ser útiles.

Hay muchos juegos que contienen una invariante que, si la conoces, te guía a ganar siempre (ver el problema 17).

Problema 17: NIM



Hay 3 filas de monedas y dos jugadores. Por turnos, cada jugador escoge una fila y de allí se lleva todas las monedas que quiera. Gana el que se lleve la última moneda. ¿Qué debes hacer para ganar?

Solución general a un problema

Cuando los datos de entrada son arbitrarios. Ejemplo: encontrar un algoritmo que resuelva cualquier problema de tipo *knapsack* (todavía no existe tal algoritmo general) o que resuelva a Pepito sus problemas con el almuerzo (esto es lo que hizo en su día al-Jwārizmī inventando el álgebra).

Solución general a todo tipo de problemas

Y con datos de entrada arbitrarios o incluso con ruido o ambiguos o mal especificados. Bueno, en eso estamos en el presente capítulo. Este es realmente el “único” problema de la IA.

Problemas clonables

También hay otra forma de clasificarlos dependiendo de si el problema es o no es clonable. La mayoría de los problemas de juegos, como el ajedrez y el Go, son clonables mientras que la mayoría de los problemas de la vida práctica no lo son.

La idea de la clonación es la siguiente: estoy jugando una partida de ajedrez contra el computador. Entonces puedo sacar un clon (una copia) del tablero y del estado de mi oponente (el computador) de tipo copia profunda⁷⁵, es decir, que no haya ningún tipo de conexión, intersección o interacción con la partida real. Sobre ese clon ensayo una jugada a ver cómo me va. Puedo sacar un número indefinidamente grande de clones, para intentar probar un gran número de jugadas, usando algún método mío o, mejor aún, algún algoritmo de IA. Cuando ya haya explorado lo suficiente y tenga claro cuál es mi mejor jugada posible, borro todos los clones y regreso a mi partida original, efectuando allí mi jugada.

En problemas de la vida real no suele ser factible clonar nada, habitualmente porque hay variables o estados desconocidos, inaccesibles o muy complejos de modelar. Por ejemplo, si compro acciones en la bolsa no hay forma de clonar el sistema para hacer ensayos no destructivos. Cualquier operación que haga sobre la bolsa tendrá consecuencias en el futuro. No hay forma de borrar ninguna acción realizada.

⁷⁵ Deep copy.

Lo mismo ocurre si intentamos enseñar algo a un niño. Cualquier error metodológico o de conocimiento que cometamos, quedará memorizado en el niño. No hay vuelta atrás. No es posible clonar para probar tácticas y seleccionar la mejor.

No obstante, hay muchos niños. Se pueden ensayar unas metodologías con unos y con otros, y comparar los resultados. Lo cual, en cierto modo, equivale al clonaje. Pero en la bolsa no está claro que se pueda hacer. Incluso aunque hay bolsas en cada país, se influyen mutuamente. No las podemos considerar clones separados. Y si dejamos pasar mucho tiempo entre un experimento y otro para que se disipen los efectos anteriores, la humanidad, la economía y las tecnologías habrán cambiado por lo que no es sensato considerar las situaciones muy separadas en el tiempo como clones.

La física experimental se basa en la suposición de que es posible clonar exactamente los experimentos (el péndulo funciona igual en todos los laboratorios de la Tierra), aunque eso solo es una suposición. Nadie sabe si el Universo guarda un estado global o si las coordenadas absolutas de un experimento afectan al resultado. A decir verdad, parece poco plausible. Sin embargo, en algunos experimentos (como cuando hay enmarañamiento cuántico⁷⁶) podría ser así.

Si recordamos el apartado “¿Qué es un simulador?”, cuando el problema es clonable, estoy creando un **tiempo interno** para realizar las simulaciones en los clones, que luego desaparece sin dejar rastro cuando selecciono la mejor acción y la ejecuto en el problema real. Allí es que el **tiempo externo** avanza un *tic* de reloj. En este sentido, los simuladores nos dan inteligencia gracias a que permiten hacer clones de los problemas y revisar así muchas hipótesis antes de tomar alguna decisión.

También podemos entender los problemas clonables como el requisito previo para que pueda trabajar (aunque sea solo en teoría) una Máquina de Turing No Determinista. Esta máquina debe fabricar clones del problema para ensayar en él todas las soluciones posibles de forma paralela y simultánea. La evolución también hace algo parecido: fabrica clones, con pequeñas mutaciones y ensaya cada uno a ver cómo se comporta solucionando un determinado problema. En este último caso la diferencia estriba en que lo que se clonan son las soluciones, en vez del problema.

Daniel Dennett (2000) nos plantea algo bastante relacionado en su libro *Tipos de Mentes*. Allí nos dice que hay cuatro tipos de mentes que, en orden progresivo de complejidad, son:

⁷⁶ *Quantum Entanglement*: cuando dos partículas interactúan, se mantienen en un estado cuántico compartido que perdura incluso aunque se separen millones de kilómetros.

- **Tipo I, mentes *darwinianas*.** Cada individuo implementa una idea, un algoritmo acerca de cómo comportarse en el mundo y que está escrito en *hardware*, fijo para siempre (habitualmente en los genes). Si el algoritmo es malo, el individuo morirá rápidamente sin dejar hijos. Cuanto mejor sea el algoritmo, el individuo tendrá más hijos que se parecerán a él (con algunas mutaciones que le darán variedad a los algoritmos). La evolución opera aquí seleccionando a los individuos con las mejores ideas. Cuando vemos animales resolver problemas, hay gente que dice que ellos no son inteligentes sino que usan su instinto. Esta es una interpretación equivocada. Lo que ocurre es que la inteligencia no está en el individuo sino en la población⁷⁷, y el aprendizaje ocurre por medio de la evolución.
- **Tipo II, mentes *skinnerianas*.** El algoritmo implantado en los genes es un poco más flexible. Si una acción recibe recompensa, el algoritmo aumenta la probabilidad de repetirla. Mientras que si recibe un castigo no mortal, el algoritmo disminuye su probabilidad. Esta realimentación conduce a un aprendizaje.
- **Tipo III, mentes *popperianas*.** Cada individuo ha construido en su cerebro un modelo del mundo, es decir, un simulador. El modelo se entrena con lo que el individuo percibe. Cuanto más inteligente sea más detalles del mundo captura en su modelo. Si es poco inteligente, el modelo podría congelarse y dejar de aprender, incluso ante eventos externos que no casen en absoluto con su modelo (sistema de creencias consolidado). Cuando el individuo tiene que tomar una decisión, primero lleva la acción que desea tomar al simulador para ver si los resultados son buenos o malos. De este modo mueren las ideas, pero no los individuos. El simulador permite hacer clones de la realidad y ello supone un gran salto en la inteligencia de los seres que lo poseen, típicamente los mamíferos.

Al respecto, un simulador es una arquitectura que permite probar modelos del mundo. El proceso de aprendizaje consiste en construir nuevos modelos y también en ajustar sus parámetros. Ser más o menos inteligente depende de tener modelos más o menos sofisticados, que van a predecir mejor o peor el futuro. Si quieres saber qué pasa cuando se sube el IVA del 15% al 18% puedes comparar los dos modelos en la figura 83. El primero es lineal, el segundo es complejo y tiene en cuenta muchos otros factores, como la tasa de desempleo, la tasa de comercio informal, la tasa de fraude y el cambio de la moneda.

⁷⁷ Como argumentaré más adelante, la inteligencia siempre es un fenómeno colectivo.

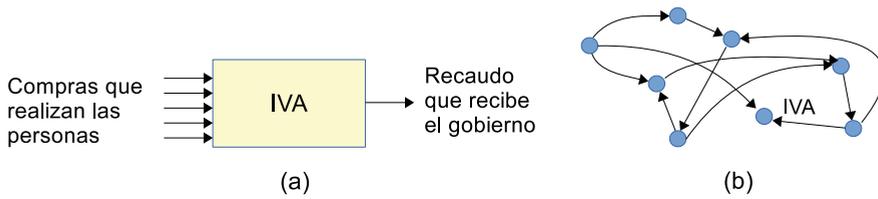


Figura 83. Dos modelos de la economía: a) lineal; b) complejo

No hace falta ir a campos tan abstrusos como la economía. En la vida cotidiana mucha gente tiene modelos erróneos. Por ejemplo, piensan que si se cae comida al suelo, disponen de 3 segundos para recogerla antes de que se contamine (Figura 84). Aquí se trata de un mero ajuste de parámetros: el primer modelo es equivocado pues supone que en el suelo hay solo unas pocas bacterias, y que ellas corren detrás de la comida. Mientras que el segundo modelo tiene un valor más realista de 10^9 bacterias por metro cuadrado, y cualquier cosa que caiga al suelo se unta inmediatamente.



Figura 84. Dos modelos de contaminación de la comida que cae al suelo

Fuente: Dibujos realizados y cedidos amablemente por Helga R. Calvo R. (2017).

Los niños pasan muchas horas entrenando sus simuladores. Es lo que llamamos jugar. Y, una vez adultos, los simuladores pueden servir para algo más, como indica Yuval Noah Harari (2016): explorar mundos imaginarios muy simples con alguna característica destacable. Son las ideologías económicas (con la idea del dinero para intercambio generalizado), religiosas (con un dios que premia y castiga), sociales de diversos tipos (con ideas como la libertad, el derecho al trabajo o la solidaridad) e incluso deportivas (clubes de fútbol). Lo interesante que propone Harari es que es muy fácil inducir a las personas a que ejecuten en su simulador estas ideologías (lo que llamamos imaginar) y, una vez que ello ocurre, los respectivos líderes consiguen controlar grandes poblaciones de humanos con muy poco esfuerzo. Los ani-

males no tienen esta capacidad por lo que leones y chimpancés están en lucha física permanente para mantener su liderazgo dentro de su manada que, a su vez, no puede ser muy grande.

De modo que el simulador que traemos de fábrica en nuestro cerebro es una característica sobresaliente que nos permite desenvolvernos mejor en el mundo, pero tiene un punto flaco y es que da acceso a otras personas para que nos manipulen. E incluso gracias a él se forman los superorganismos sociales. De alguna forma es también el policía interno que nos vigila, una vez que dejamos que se ejecute en él una cierta ideología, para que nos alineemos con lo que la sociedad espera de nosotros.

Por último, esta idea de Harari de que el *hackeo* del simulador cerebral permite crear ficciones que agrupen a los humanos en superorganismos, encaja muy bien con el concepto de memes ideado por Richard Dawkins (1994) y desarrollado por tantos otros científicos, entre los que destacan Daniel Dennett, Susan Blackmore y Robert Aunger (2002). Los memes son unidades básicas de información tales como teorías científicas, mitos religiosos, modas, temas musicales, obras de arte y prácticamente cualquier idea que podamos pensar. Los memes son los individuos de un proceso evolutivo, pues cumplen con las cuatro condiciones requeridas para que haya evolución. Hay una población de memes, que sacan copias de sí mismos saltando de un cerebro humano a otro. Siempre habrá variabilidad, pues las ideas se transmiten con errores o se combinan unas con otras. En este escenario, los cerebros humanos serían entonces las máquinas de reproducción de los memes. Lo que Harari aporta muy brillantemente es el mecanismo de presión selectiva, que nunca había sido explicado de forma convincente por los otros autores. Los memes que más se reproducen son los que fabrican las ficciones que tienen más éxito en el cerebro de los humanos, generando grupos sociales cohesivos que propagan esas mismas ideas.

- **Tipo IV, mentes gregorianas.** La mente anterior puede predecir el futuro, mientras que esta también puede modificarlo, pues es capaz de construir y usar herramientas (desde martillos hasta computadores) incluyendo un lenguaje para comunicarse con otros individuos. Aquí Dennett comete un error de categorización, pues este tipo de mente existe, efectivamente, pero no es la culminación de las tres anteriores sino que corre paralelo con ellas. La razón de ello es que construir herramientas no es tanto una habilidad cerebral, que por supuesto lo es, sino sobre todo la disponibilidad de tener una mano o algo similar, capaz de manipular el mundo. No es un asunto de potencia de cómputo sino de

versatilidad en el sistema de entrada-salida, usando la jerga de los computadores. Este nivel no lo pudieron alcanzar los delfines, a pesar de que son muy inteligentes y seguramente tienen una mente *popperiana*. Sí lo pudieron alcanzar los humanos, los simios en general y los elefantes con su trompa. Les queda muy difícil a los caballos y las vacas. Y de la misma manera que la Máquina de Turing Universal logra computación completa, la mano logra entrada-salida completa, ya que puede construir cualquier mecanismo incluyendo microscopios para ver lo muy pequeño, telescopios para ver lo muy grande, sensores para recibir información de entrada que antes no era accesible (por medio de infrarrojo, ultravioleta, rayos X, vibraciones o detección de productos químicos), e incluso otras máquinas para fabricar otros mecanismos imposibles de hacer a mano, donde también se incluyen máquinas de fabricación universales como las impresoras 3D, y máquinas de cómputo.

De hecho, hay muchos animales que construyen herramientas, aunque la mayoría lo hace porque están impresas en sus genes o, como dice Dawkins (1982), forman parte de su fenotipo extendido. Ejemplos son los termiteros, las telas de arañas, las madrigueras y los nidos. En la figura 85 podemos ver el nido de un tipo de araña, cuyos genes construyeron la tapa. Esta tapa se podría considerar como una parte de su cuerpo, a la que no hay que alimentar con oxígeno y azúcar. Pero, al fin y al cabo, también es una herramienta que le sirve para que no entre la lluvia o enemigos a su nido. Otros ejemplos de herramientas, esta vez no impresas en genes sino transmitidas culturalmente, son las piedras para abrir semillas que descubren algunos simios. El lenguaje de cada especie también puede considerarse una de estas herramientas, y parece que tiene las dos componentes: una genética y otra socialmente aprendida (Pinker, 1994).



Figura 85. Nidos de araña con tapa, en la Chapada Diamantina, Brasil

¿CÓMO RAZONAN LOS ROBOTS?

Si quieres profundizar en las últimas tendencias, deberías leer el libro de Pedro Domingos, *Machine Learning* (2015). Para él, la inteligencia también es un asunto de predicción, con la diferencia de que, en aprendizaje de máquina, los algoritmos ya no los escriben los humanos, sino que el computador los aprende por sí mismo. Pedro clasifica estos algoritmos en cinco técnicas: razonamiento simbólico, redes neuronales, computación evolutiva, razonamiento *bayesiano* y razonamiento por analogías. Veremos unas nociones someras de cada una de ellas.

Razonamiento simbólico. Es puramente lógica matemática. En sus inicios se basaba en los lenguajes *LISP* y *PROLOG*, que se pensaba iban a ser la base de la inteligencia artificial. Se implementa con árboles de decisión o con sistemas de reglas clasificatorias de tipo *IF-THEN-ELSE*. Se necesita un experto que programe las reglas adecuadamente, aunque en el libro anterior también vimos que se puede hacer evolutivamente. Al ser exacto y determinista se le aplica el teorema de Gödel, es decir, no puede ser creativo, no puede inventar nuevos espacios y conceptos. Pero sí puede explorar sistemáticamente todo un espacio ya existente. Tampoco puede lidiar con incertidumbres, probabilidades, especificaciones contradictorias ni ambigüedades, de modo que está muy poco adaptado a los problemas reales que tenemos que solucionar los humanos. Esto quizás lo entienden poco los filósofos: el problema del marco⁷⁸, la objeción de Ada Lovelace⁷⁹ y otros obstáculos a la inteligencia artificial solo aparecen en este tipo de algoritmos.

Para operar requiere únicamente disponer de una Máquina de Turing Universal. Y ya hemos visto en el libro anterior que estas máquinas son de muy baja complejidad. Esta es la única técnica que ofrece resultados exactos, lo que, paradójicamente, es su principal desventaja ya que ello implica diversas limitaciones que la impiden alcanzar la verdadera inteligencia, aunque puede ser un apoyo para las otras técnicas.

Redes neuronales. Son un modelo simplificado de la arquitectura *hardware* del cerebro humano. Cada neurona se puede considerar un nodo donde se efectúa un cómputo a partir de los valores que llegan a través de sus arcos de entrada, que simulan las dendritas de una neurona real. El resulta-

78 Diferenciar lo relevante de lo irrelevante para resolver un problema de forma eficiente. Por ejemplo, si tengo setas de varios colores y quiero saber cuántas son, ¿el color es relevante? ¿Y si quiero comerlas sin envenenarme? Cualquier problema tiene miles de aspectos irrelevantes, y descartarlos sistemáticamente puede consumir demasiado tiempo.

79 Los computadores no son inteligentes. Son inteligentes los humanos que los programan.

do se ofrece por un arco de salida, que simula el axón. Los arcos de entrada están etiquetados con unos pesos fijos, y están conectados a las salidas de otras neuronas. Cada neurona calcula el sumatorio de sus entradas multiplicadas por los respectivos pesos, a lo que aplica una función de activación y si el resultado supera un umbral, la salida se activa.

Las redes neuronales se suelen diseñar con varias capas de neuronas, típicamente tres. La primera capa recibe las entradas del problema, la segunda capa hace un procesamiento, y la tercera capa genera las salidas, o sea, los resultados.

Hay una primera fase de entrenamiento supervisado, donde se le presentan a la red neuronal problemas con sus soluciones. La red entonces ajusta los pesos de sus conexiones entre neuronas, usando algoritmos como el *backpropagation*. También hay otras redes capaces de hacer aprendizaje no supervisado, donde se les presenta problemas de entrada, pero no se les dice cuál es la solución correcta.

La principal limitación de estas redes es que los algoritmos de aprendizaje donde calculan los pesos, suelen ser lentos y no convergen cuando hay más de tres capas. Sin embargo, se ha inventado un nuevo tipo de arquitectura que solventa este problema: el *autoencoder*. Consiste en una red de tres capas que intenta predecir sus entradas, es decir, la tercera capa está conectada a la primera. El entrenamiento es no supervisado y cuando ya están entrenadas, la segunda capa contiene aproximadamente la misma información que la entrada pero comprimida. Se puede forzar que haya compresión poniendo menos neuronas en la segunda capa que en la primera. Recordemos que comprimir es una forma de aprender o, si se quiere, de generalizar. Por ejemplo, si en la primera capa hemos puesto los *pixeles* de la foto de un gato, y hemos entrenado la red para que en la última capa aparezca más o menos esa misma foto, entonces en la capa intermedia habrá algo esencial de la foto a partir de lo cual se puede reconstruir el original.

Otra forma más sofisticada de forzar la compresión es poniendo muchas más neuronas en la segunda capa que en la primera pero obligando a que la mayoría de ellas no se activen. Con ello se consigue que el entrenamiento sea más fácil y, a la vez, que el resultado sea un conjunto de características separadas de la entrada. Ello da lugar a que la entrada quede codificada de una forma similar a una matriz dispersa, también usada por Hawkins en Numenta (2005). Es interesante destacar que el cerebro humano realiza algo similar, aunque de forma distribuida.

Nada impide apilar *autoencoders*. Es decir, usar un segundo *autoencoder* para comprimir aún más la información que entrega el primero. O, dicho de otra manera, para asociar de otra forma las características que entrega el

primero. Y luego utilizar un tercero para hacer lo mismo con el segundo, y así sucesivamente. De este modo, si a la entrada ponemos los *pixeles* de la foto de un gato, quizás el primer *autoencoder* extraiga características básicas de la imagen, como segmentos, zonas de color similar y así. El segundo *autoencoder* puede extraer siluetas, separar el primer plano del fondo y otras operaciones resultantes de analizar las características entregadas por el primero. El tercer *autoencoder* quizás sea capaz de reconocer la forma del gato analizando los datos entregados por el segundo.

Hay un éxito impresionante con las nuevas arquitecturas de redes neuronales, gracias a los *autoencoders* y al gran número de capas que se pueden poner u entrenar. Y eso es lo que actualmente se llama *Deep Learning* (Géron, 2017). Estas redes son las que ganaron el campeonato mundial de Go, las que reconocen gatos en videos, las que aprenden a jugar y ganar cualquier juego de computador por sí mismas.

La complejidad de estas redes neuronales es bastante baja pues no necesitan computación completa ni para ejecutar la red ni para entrenarla, y estructuralmente son solo un grafo dirigido.

Algoritmos evolutivos. Dado un problema, el programador crea un patrón general para la solución, llamado cromosoma. Ese patrón es un contenedor de datos y programas que se requieran para calcular la solución. Inicialmente los valores concretos de los datos y programas son desconocidos, y lo que se busca es averiguar cuáles son. Se parte de una población de cromosomas generados al azar que se hacen evolucionar, para lo cual se calcula la aptitud de cada uno —es decir, lo bueno o malo que es— seleccionándose los mejores con mayor probabilidad, para reproducirlos. La reproducción puede hacerse de muchas formas, pero lo que se busca es que los hijos se parezcan a los padres, pero que no sean completamente idénticos. Eso se logra con la mutación, que consiste en hacer un pequeño cambio a un padre, y con el cruce, que consiste en tomar una parte de un cromosoma y otra parte de otro, combinándolos entre sí. Los hijos se introducen en la población y opcionalmente hay otro proceso de selección para matar algunos cromosomas. Así se completa una generación. El algoritmo corre por muchas generaciones hasta que aparezca alguna solución aceptable.

Puede haber cromosomas que mantengan hipótesis contradictorias, lo cual es bueno, pues se deja que la evolución decida cuál es la mejor. Y la solución mejor encontrada seguramente no es la mejor posible, pues conforme se acercan al óptimo la presión selectiva disminuye.

Es un algoritmo lento, pero muy general, pues se puede aplicar a cualquier tipo de problema, incluso los que están mal definidos. Se puede considerar que cada cromosoma es una hipótesis respecto a la solución de un

problema, y que el proceso evolutivo genera nuevas hipótesis y selecciona las que están mejor adaptadas para solucionar el problema.

La principal limitación es que, una vez definido el cromosoma, el espacio de búsqueda queda allí limitado, es decir, la creatividad del algoritmo se desarrolla dentro de ese espacio, pero no puede crear espacios nuevos. Hemos visto muchas variantes de estos algoritmos en el libro anterior.

Estos algoritmos están inspirados en la evolución biológica que descubrió Darwin. Pero en el cerebro humano ocurren también procesos similares para la toma de decisiones: consideración de un conjunto de hipótesis, fabricación de otras por medio de pequeñas variantes, selección de las mejores, iterar unas cuantas veces y al final elegir la mejor.

La complejidad de un algoritmo evolutivo es prácticamente la misma que la reproducción. Si un objeto sabe reproducirse, su evolución es prácticamente inevitable.

Razonamiento bayesiano. Se tiene un conjunto de hipótesis mutuamente excluyentes y se tienen creencias sobre lo bien que funciona cada una de ellas en un problema específico. Las creencias se modelan con probabilidades, también llamadas probabilidades *a priori*. Habitualmente se escoge la hipótesis en la que se crea más (la que tenga la mayor probabilidad de éxito) para tratar de solucionar el problema. Una vez que sabemos la solución al problema, tanto si hemos acertado como si no, actualizamos nuestras probabilidades *a priori*, usando el famoso teorema de Bayes.

Razonamiento por analogías. Aquí Pedro agrupa varios algoritmos no completamente relacionados: la búsqueda del vecino más cercano, las máquinas de soporte vectorial y el razonamiento analógico completo. El objetivo de razonar por analogías es poder llegar a conclusiones aproximadas aunque casi no haya datos de entrenamiento. Basta con un ejemplo sobre un tema específico, para poder trabajar con esta técnica y comenzar a obtener resultados, así sean aproximados.

Hay también combinaciones de todo ello, como la lógica difusa, lógica multivariante y la lógica temporal, que están basadas en el razonamiento simbólico, pero cada proposición tiene un *tag* añadido que expresa algo nuevo que con la lógica formal no se puede, como por ejemplo, la probabilidad de que la proposición sea válida, el intervalo de tiempo de validez, el grado de verdad, y cosas similares. Y se han unificado varios marcos teóricos, aunque no todavía los cinco.

Pero eso no significa que aquí terminen todas las posibilidades. Quedan muchos temas por explorar, y para saber cuáles, nada mejor que entender en qué se basan los razonamientos humanos.

¿CÓMO RAZONAMOS LOS HUMANOS?

Para poder fabricar inteligencia artificial es una buena idea fijarnos cómo funciona la inteligencia humana. Es interesante conocer el detalle de las conexiones neuronales, pero no con el objetivo de imitarlas como intentan algunos proyectos de investigación. Eso no tiene mucho sentido. El objetivo será sacar modelos que capturen lo importante de la inteligencia, eliminando los detalles accesorios. Recordemos que, inspirados en las aves, hemos fabricado aviones, pero ninguno necesita de detalles como las plumas.

Vuelvo a insistir en que la inteligencia biológica es el resultado de un proceso evolutivo, y que el problema que trata de resolver es sobrevivir y reproducirse dentro de grupos sociales, y no calcular integrales triples. Eso puede llegar después, como un efecto secundario.

¿Cómo apareció por evolución la inteligencia humana? Hay varias hipótesis y me centraré en la más ampliamente aceptada. Como argumenta Dawkins (1994), la cooperación dentro de un grupo no puede tener un sustento genético. Y la cooperación es lo mínimo indispensable para que el grupo exista, porque de lo contrario le iría mejor a cada uno viviendo aislado de los demás. Sin embargo, es un hecho que existe la cooperación en los grupos de primates e incluso de otros mamíferos. Cuando un grupo coopera para conseguir un objetivo común, es inevitable que alguien trate de fingir cooperación sin realmente trabajar para el grupo, obteniendo de todos modos los beneficios de pertenecer a él. Como bien explica Pinker (1997), ello debió crear una carrera de armamentos entre cerebros que se hacían cada vez más grandes tratando de detectar los engaños de los otros miembros de la sociedad para evitar abusos, a la vez que intentando engañar a los otros para su propio beneficio. Parece ser que esa es la causa de que el cerebro del *Australopithecus Afarensis*, de 400 cm³ se agrandara en 4 millones de años hasta llegar a nuestros 1500 cm³ en promedio, a la vez que aumentaba el tamaño de la sociedad en la que vivíamos.

En esa carrera de armamentos, para lograr engañar a los demás, puede ser una buena táctica engañarse primero a sí mismo, como propone Trivers (2011), de modo que el otro no pueda leer en el rostro alguna incongruencia y así no dejar ningún resquicio a la duda. En general no es razonable esperar que un animal quiera engañarse a sí mismo, pues al alejarse de comportamientos óptimos podría poner en riesgo su vida. Sin embargo, en ambientes sociales es distinto. Juan Camilo Ramírez fue un estudiante asiduo a EVALAB, aunque su trabajo de grado finalmente no lo realizó aquí, y ha seguido trabajando en temas evolutivos durante sus estudios de doctorado. En Ramírez y Marshall (2015) nos muestra un modelo donde el autoengaño

evoluciona y prospera. Como efecto secundario, un cerebro tan grande diseñado para predecir y anticipar el comportamiento de sus congéneres puede servir para muchas otras cosas. Eso es lo que se llama *exaptación*. El cerebro humano no fue diseñado para las matemáticas y la ingeniería (los computadores son más precisos y más rápidos), pero puede servir para ello. Y es importante saber que, debido a su origen cooperativo, la forma de razonar de los humanos está siempre impregnada de un tinte social (ver problema 18).

Problema 18: Detectar engaños

a) Nos muestran las siguientes cuatro cartas que tienen una letra por un lado y un número por el otro y nos aseguran que “*si una carta tiene la letra D en un lado, entonces tiene el número 3 en el otro*”:



¿Qué cartas hay que girar para saber si esa afirmación es verdadera?

===

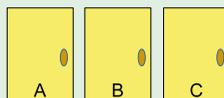
b) Supongamos que usted es un camarero en un bar y está haciendo cumplir la regla “*si una persona está bebiendo cerveza, ella debe haber cumplido 18 años*”. Usted puede preguntar a las personas que quiera qué están bebiendo o cuál es su edad. En la barra hay cuatro personas: un bebedor de cerveza, un bebedor de naranja, una persona de 25 años y una persona de 16 años.

¿Qué debe preguntar y a quién?

A la mayoría de la gente los problemas abstractos no se le dan bien, pero los sociales sí. Por eso las matemáticas suelen ser difíciles para casi todos los jóvenes, aun cuando sean muy inteligentes. Muchas ramas de las matemáticas tienen un fundamento básico que puede tener un origen social, como contar objetos para intercambiarlos o repartirlos. Pero otras son demasiado abstractas y hay que reconocer el gran avance que supuso su mera formulación. Un ejemplo es el teorema de Bayes, y en general, toda la estadística (ver el problema 19).

Problema 19: Las tres puertas

Estamos en un concurso de televisión, donde nos muestran 3 puertas. La mecánica del concurso es la siguiente: detrás de una de ellas está el premio, mientras que en las otras no hay nada. El presentador, quien sabe perfectamente en qué puerta está el premio, nos pide que elijamos una, pero todavía no la abre. Para aumentar la emoción, el presentador abre otra puerta, y nos muestra que está vacía. Por último, el presentador nos pregunta si queremos permanecer con la puerta elegida en un principio o deseamos cambiarnos a la tercera puerta que aún permanece cerrada.



¿Cuál es la mejor estrategia, permanecer o cambiar?

Teniendo en cuenta que inevitablemente nuestro razonamiento lleva incorporado un tinte social, que incluso puede verse como una especie de defecto de fábrica, podemos presentar una primera clasificación de cómo resolvemos problemas los humanos:

- **Fe.** Suponer que elegir cierta acción nos va a traer más beneficios, pero realmente no tenemos ninguna prueba, certeza, indicador ni experiencia de ello. Esto es equivalente a una variable estocástica. No es recomendable actuar por fe ni individualmente ni colectivamente.
- **Intuición.** Tenemos pruebas, experiencia e indicadores de nuestras ideas, pero no las hemos sabido formalizar. Hay bastante probabilidad de éxito y es recomendable seguir intuiciones a nivel individual, pero no podemos comunicar de manera efectiva este tipo de razonamientos a un colectivo de personas.
- **Razonamiento estructurado.** Hemos reflexionado sobre nuestras intuiciones y hemos logrado verbalizarlas y modelarlas en un proceso lógico inductivo, deductivo u otro. Esta es la forma de inteligencia más fuerte, en el sentido de que podemos someter a pruebas el modelo. Además, estos modelos son comunicables al colectivo de personas, por ejemplo, por medio de la universidad.

Pongamos ejemplos de los tres casos:

- A Juan le gusta el fútbol, pero a su equipo del alma le está yendo muy mal. Han perdido los 10 últimos juegos por su pésimo estilo, el presidente está en la cárcel por una cuestión de sobornos a árbitros, la hinchada no asiste ya al estadio, tienen muchas deudas, el entrenador acaba de despedirse porque no le están pagando, y hay varios jugadores lesionados. A pesar de ello, Juan tiene fe en que va a ganar el partido de esta noche. Eso es la fe. Irracional. Todos los indicadores en contra, pero Juan tiene fe. La fe a veces se basa en la tradición, es decir, seguir en lo que siempre se ha hecho o se ha visto hacer a los seres cercanos. A veces se confunde con el deseo. Pero definitivamente es un comportamiento humano. Es irracional pero vemos cómo a diario los humanos lo usamos para tomar decisiones. A veces incluso graves decisiones. De modo que hay que considerarlo como parte de la inteligencia humana. Una forma de modelarlo es como una variable estocástica, que toma un primer valor al azar, y ya no lo cambia jamás. Probablemente nadie quiera que la fe se incluya dentro de un *software* de inteligencia artificial, pero ese es otro asunto.
- Luisa acaba de despertar de un letargo de 50 años en que estuvo bajo criogenia. Y mira sorprendida un televisor bastante extraño, pero

televisor, donde va a comenzar un partido internacional entre Estados Unidos y Albania, de un juego llamado *calamunca*, completamente nuevo para ella, pero que causa furor por estos días. Ella, en los breves minutos que lleva despierta, intuye que va a ganar Estados Unidos. Y acierta. Más adelante, si le dan suficiente tiempo para reflexionar sobre ello, Luisa nos dirá que las reglas del juego le son desconocidas y que le costó un buen rato medio entender de qué se trataba, pues ni siquiera había una pelota rodando. Pudo identificar dos equipos por el color de los uniformes, y un montón de zonas en el suelo, que también tenían colores diversos. Y por los gritos de los aficionados sabía que estaba ocurriendo una jugada importante. Pero los movimientos de los jugadores eran completamente incomprensibles para ella. Sin embargo, sabía qué países eran Estados Unidos y Albania hace 50 años. El primero grande y rico, el segundo pequeño y pobre. Esas cosas no suelen cambiar rápidamente. Y en un país grande y rico hay más posibilidades de invertir en deporte, y hay más gente que haga deporte, y por ello hay más probabilidad de encontrar mejores deportistas. Además, en la mayoría de deportes que ella conoce, el primer país ganaba con holgura al segundo hace 50 años. Entonces basó su respuesta en esta información. Luisa está verbalizando ahora todo esto, pero en el momento de hacer la predicción, simplemente la hizo, sin razonarla detalladamente como ahora. Eso es lo que llamamos intuición. Tenemos un montón de conocimientos almacenados que usamos cuando las circunstancias no nos dan tiempo a hacer un razonamiento detallado. Y en muchos casos, acertamos. La intuición sí sirve, especialmente cuando hay que tomar una decisión bajo la presión de un tiempo escaso. Obviamente unos humanos tienen mejor intuición que otros. La intuición forma parte de la inteligencia humana. Y sería muy deseable que la inteligencia artificial también usara la intuición cuando opere con restricciones de tiempo. Pero no es comunicable, no se puede enseñar en las universidades, depende de un montón de experiencias personales.

- Luisa lleva ya cinco años en su nueva realidad, y también se aficionó al *calamunca*. En su país hay muchos equipos y también hay un sistema de apuestas con el que ella está ganando bastante dinero. Nos cuenta que poco a poco fue extrayendo un modelo de cada jugador, de todos los equipos. El modelo incluye la velocidad, los reflejos, la coordinación con otros y la visión lateral, que son los factores relevantes del juego. Una vez que sabe las alineaciones de cada equipo, corre muchas veces el modelo añadiendo ruido en todos los parámetros para verificar si emerge una solución robusta. La apuesta la hace sobre

los equipos ganadores de su modelo, y el valor de la apuesta depende de la dispersión de los resultados. Crear un modelo de una situación real, ajustar correctamente sus parámetros y tomar decisiones sobre el mundo real basadas en los resultados de la ejecución del modelo constituyen un ejercicio de inteligencia de un nivel más alto. No es la única forma de inteligencia, pues también se pueden tomar decisiones basadas en métodos deductivos, inductivos y otros.

Hemos visto tres niveles de inteligencia. ¿Puede haber más? En principio, nada lo impide, pero no tenemos más pistas al respecto. Probablemente surjan como combinatoria de los métodos ya enumerados. Quizás alguien descubra métodos nuevos. Y a partir de aquí solo podemos elaborar más el tercero, el razonamiento estructurado, que a su vez tiene varias formas:

- **Deducción.** Es el método más exacto, donde se parte de verdades generales para demostrar verdades particulares. El ejemplo más obvio es el *modus ponens*, que podemos ilustrar con un ejemplo: *si este jarrón se cae, se romperá. El jarrón se ha caído. Por tanto, el jarrón se ha roto.*
- **Inducción.** Aquí se hace al revés. Se parte de verdades particulares y se llega a una verdad general. Es un método exacto en matemáticas, y es convincente, pero no exacto, en las ciencias. Para hacernos una idea de por qué jamás la inducción será exacta⁸⁰ en ciencias, nada mejor que un chiste: *un matemático, un físico y un ingeniero viajan juntos por primera vez por Australia. Van en un autobús y divisan un cisne de por allí. El ingeniero dice: ¡mirad, en Australia todos los cisnes son negros! El físico le replica: no. Lo que sabemos es que en Australia hay un cisne negro. Y el matemático sentencia: no, no, no. Lo único que sabemos es que en Australia hay un cisne que es negro por un lado.* Es decir, nunca hay seguridad de que hayas logrado capturar la esencia de un fenómeno, y que el sol haya salido mil veces seguidas por el este no te da garantías de que mañana vuelva a hacerlo. También se le llama a ello el razonamiento (erróneo) del pavo que, el día antes de navidad piensa: *estos humanos son muy amables pues han estado alimentándome y cuidándome durante muchos meses. Nada malo me ocurrirá con ellos.*
- **Reducción al absurdo.** Se supone la hipótesis falsa y se razona con ella hasta llegar a una contradicción, lo que significa que la hipótesis no puede ser falsa. Un ejemplo en matemáticas sería la demostración de Euclides que dice que los primos son infinitos: primero suponemos que los primos son finitos $\{p_1, p_2, \dots, p_N\}$. Los multiplicamos todos y le

80 El filósofo David Hume alertaba de ello en el siglo XVIII.

sumamos 1 al resultado, obteniéndose un número $q=p_1 \cdot p_2 \cdot \dots \cdot p_N + 1$. Este nuevo número q es mayor que todos los demás y no puede ser primo debido a la suposición inicial. Por tanto es no-primo, y eso significa que existe un número primo p_x que es divisor de q . Desde luego, ese primo p_x es divisor de $p_1 \cdot p_2 \cdot \dots \cdot p_N$ y dado que $q=p_1 \cdot p_2 \cdot \dots \cdot p_N + 1$ entonces también se requiere que p_x sea divisor de 1 , lo cual es imposible. Por tanto, la suposición inicial es falsa, y los primos son infinitos.

- **Diagonalización.** Sirve para hacer demostraciones con conjuntos de infinitos elementos, típicamente para demostrar que la cardinalidad de dos conjuntos es distinta, aunque en ambos sea infinita. Ello creó la matemática de los transfinitos, que sirvió para demostrar que hay más números reales que naturales. Y también es el núcleo de la demostración de Gödel y del problema de la parada de Turing.
- **Mutación.** Es decir, prueba y error, que es lo más básico. Con este método no se pueden hacer razonamientos, pero sirve para descubrir verdades probables (que después habrá que demostrar con otro método).
- **Amplificación del absurdo.** Es, básicamente, el método *socrático*: hablar con una persona de un tema sobre el que mantiene una opinión absurda, animarla a que lo desarrolle, hacerle preguntas, para finalmente que ella misma se dé cuenta que llega a una contradicción. Es especialmente interesante y fácil de practicar con niños. Un computador no necesitaría de este método, pues al llegar a una contradicción la detectaría rápidamente. Pero los humanos somos propensos a vivir mucho tiempo con disonancias cognitivas sin darnos cuenta. El propósito de este método es entonces amplificarlas hasta que sea insostenible mantenerlas y fuercen a un cambio de ideas.
- **Aumentar el número de dimensiones.** Al aparecer una contradicción por alguno de los métodos anteriores, se debe reformular todo. Para ello hay que ser creativos (usando mutaciones y analogías) buscando enmarcar el problema en un sistema de hipótesis más amplio, con más dimensiones (grados de libertad), que evite la contradicción. Este es el método típico de la física fundamental para crear nuevas teorías. Viene sugerido por la metáfora de que hay nudos hechos con cuerdas vulgares que no pueden desatarse en nuestro mundo de tres dimensiones, pero si pudiéramos sumergirlos en un mundo 4D entonces sí se podrían desatar. Un ejemplo real es que como la física de Newton entra en contradicción cuando se encuentra que la velocidad de la luz es constante aun cuando el observador esté viajando a cualquier velocidad, entonces se relaja la suposición de que el espacio y tiempo son sistemas de referencia fijos y se les permite estirarse y encogerse,

dando lugar a la relatividad especial. Por así decir, se crea una nueva variable que antes no existía: el factor que dice cuánto se encoge el espacio y cuánto se estira el tiempo.

- **Búsqueda del movimiento mínimo.** De manera análoga a cómo calcular el vector unitario ortogonal en un espacio matemático nos permite llegar cómodamente a cualquier punto de él, averiguar cuál es el movimiento mínimo nos permite alcanzar cualquier estado de un problema. Por ejemplo, en el cubo de Rubik el movimiento mínimo permuta 3 casillas cuadradas de sitio. Es imposible cambiar solo 2 o 1. Sabiendo que ese es el movimiento mínimo se ha creado después una gramática que combina este movimiento respecto a los 3 ejes, para generar cualquier estado posible del cubo. A partir de allí podemos operar con el lenguaje definido, en vez de con el objeto físico, y ello permite emplear técnicas clásicas de búsqueda de la solución, como los algoritmos evolutivos.



Figura 86. Puzzles que consisten en separar las piezas constituyentes. Y, también, en volverlos a armar como estaban en un principio

En problemas como el de la figura 86 se comienza haciendo movimientos al azar (mutaciones) que nos permiten explorar el espacio de búsqueda, no tanto para encontrar la solución, sino los movimientos mínimos que cambian el objeto a otros estados. Con esos movimientos y con los estados identificados se puede construir un

lenguaje (típicamente una máquina de estados finitos) con la que realizar ya la búsqueda de la solución.

- **Analogía, metáfora o isomorfismo.** Es el más potente hasta ahora, y se usa casi sin darnos cuenta. Consiste en crear un paralelismo entre dos áreas distintas de conocimiento, de modo que los razonamientos que se hagan en una, se pueden trasladar a conclusiones en la otra. Uno puede pensar que la metáfora es una figura retórica confinada al ámbito literario. Nada más lejos de la realidad. La ciencia está plagada de metáforas e incluso el immaculado edificio de las matemáticas las usa sin reparo. Por ejemplo, Kolmogorov (1973) emplea un lenguaje muy bonito, muy matemático, conciso, limpio, a la vez que, con frecuencia usa expresiones aparentemente poco rigurosas: “surge de modo natural este nuevo concepto”, “es muy natural considerar este nuevo concepto así”, “análogamente”, y expresiones similares. Hay que afrontarlo. La metáfora es una forma de razonamiento muy imbricada en el pensamiento humano y la única que nos permite dar el salto entre el mundo de las ideas y el mundo real. Cuando pasamos de sumar dos manzanas con tres manzanas, a sumar $2+3$, estamos haciendo uno de estos saltos. Hofstadter (1995) es el principal investigador que hace énfasis en que realmente este es el método que más usamos los humanos. En la figura 87 tenemos un ejemplo de muchos *puzzles* que realmente son el mismo, si uno sabe buscar el isomorfismo adecuado.



Figura 87. Puzzles isomorfos

- **Búsqueda de los límites.** Cuando no se tiene la menor idea de cómo solucionar un problema, es una buena estrategia formular los aspectos teóricos que impiden encontrar la solución. Por eso es que hay un capítulo en el libro anterior que explica cuáles son los límites fundamentales de la computación, que nos impiden llegar a la verdadera inteligencia artificial. Una vez que se conocen los límites se puede intentar ver en qué circunstancias no son aplicables. O se puede intentar formular el problema en otros términos, donde desaparezcan las premisas que soportan los límites. En el capítulo siguiente sobre la consciencia se verá un ejemplo de ello: una vez que aceptamos que ningún proceso físico, químico o biológico tiene la capacidad causal para generar consciencia, ello significa que hay que buscar otro tipo de procesos, concretamente, computacionales.

El razonamiento estructurado humano está parcialmente cubierto por las técnicas de *Machine Learning* que explicamos en el apartado anterior. Pero no completamente, y aquí queda todavía mucho trabajo por hacer, explorando las técnicas recién enumeradas. Y quizás haya más.

Además, esta es una clasificación algo engañosa. Los humanos solo razonamos así cuando nos enfrentamos a un examen o algún problema formal o en ámbitos académicos como la escritura de un artículo o dar clases. En la vida diaria nadie usa el método deductivo para saber si puede cruzar sin peligro una calle, ni usa la reducción al absurdo para elegir pareja. Está bien documentado que los humanos tomamos decisiones basados en las emociones (Ariely, 2008) y luego racionalizamos el proceso. Y que cuando en teoría de juegos se habla de jugadores racionales, no entran allí los humanos que son capaces de hacerse daño a sí mismos con tal de llevar a cabo una venganza, porque eso no es racional. Entonces, todos estos métodos tienen que ver con un ideal de racionalidad que perseguimos para la inteligencia artificial.

El filósofo David Hume (Quinton, 1999) proponía que la razón es, y debe ser, esclava de las pasiones, como la prudencia, el amor o la envidia. La razón no puede tomar decisiones por sí misma, sino que cada persona tiene pasiones y ello le lleva a desear hacer cosas. De este modo, la razón le ayuda a encontrar la mejor forma de hacerlas. Esto ha dado lugar a una nueva área de investigación, los ordenadores emocionales, uno de cuyos primeros autores fue Picard (1998).

INTELIGENCIA COLECTIVA

Toda inteligencia es colectiva,
ÁNGEL E. GARCÍA BAÑOS

Últimamente han aparecido varios libros muy interesantes sobre inteligencia colectiva, con ejemplos de cómo un conjunto de agentes (personas, hormigas, bacterias, hongo del fango en laberintos...) pueden resolver problemas que ninguno de ellos por separado podría.

En cuanto reflexionamos sobre ello resulta que toda inteligencia es colectiva, pues cuando decimos que el ser humano es inteligente, habitualmente lo que queremos resaltar son los logros de toda la humanidad, que se basan principalmente en un sistema educativo de replicación de conocimientos. Si yo puedo desarrollar un *software* interesante es porque millones de personas antes que yo han trabajado, ideado y proporcionado un conjunto de teorías y de herramientas físicas (incluyendo los computadores) que me permitieron dar mi pequeño pasito hacia adelante.

Pero si nos centramos en un solo ser humano realizando una tarea para la que se requiere inteligencia, debemos percatarnos que quien realmente está pensando es un colectivo de neuronas.

Y probablemente por ello es que un *software* monolítico diseñado con una intencionalidad concreta no nos creemos que pueda llegar a ser inteligente pues no supera la objeción de Ada Lovelace: ese *software* hace ciegamente lo que el programador le pidió que hiciera, por lo que la inteligencia reside en el programador, no en el *software*.

La cosa cambia si diseñamos un *software* formado por un colectivo de programas, de cuya interacción surgen el pensamiento o las acciones que vamos a ver que expresa hacia el mundo. Ese *software* sí nos puede sorprender y quizás podamos atribuirle algún tipo de inteligencia. La sorpresa (o el no-determinismo) viene de la interacción no coordinada entre muchos algoritmos.

Al respecto los algoritmos evolutivos son una buena opción porque están constituidos por una población de entes (estructuras de datos o programas) que compiten entre sí y se reproducen. Solo uno de los entes se expresará hacia el mundo, solo uno de ellos propondrá su solución al problema que le planteemos. Pero no es posible predecir cual, especialmente porque cada programa habrá sufrido modificaciones aleatorias por interacción con otros programas (mutación y cruce), de modo que ni siquiera el programador podrá saber de antemano cuál va a ser la solución que proponga el algoritmo evolutivo.

Y esto tiene un claro paralelismo con la forma como funcionamos los humanos. Ante un problema, es común que debatamos internamente entre varias soluciones posibles. Es posible que ideemos otras nuevas (mutaciones) o combinemos varias (cruce). Finalmente una de esas ideas internas la llevaremos a cabo, y de lo malo o bueno de su resultado nos quedará una experiencia que modificará la forma como realicemos este proceso en un futuro. Incluso es posible que hagamos algo un día y cambiemos de idea al día siguiente, lo que quiere decir que la población de ideas no desaparece incluso aunque haya una vencedora, sino que continúan en una lucha interna por salir a expresarse al exterior.

En el libro *The wisdom of crowds* (2014), James Surowiecki nos cuenta cuándo comenzó a pensarse en inteligencias colectivas. Parece ser que el estadístico Galton estaba en un mercado de ganado y presencié un concurso: había un buey y la gente debía tratar de adivinar cuanto pesaba, rellenando un papelito con sus datos. Al finalizar el concurso pesaron el buey, miraron todos los papelitos y dieron un premio al que más se acercó en su pronóstico. Galton pidió los papelitos y sacó el promedio de todos ellos. Este resultado se acercó mucho más al peso real del buey. Con este ejemplo se muestra que el colectivo sabe más que el mayor de los expertos.

¿O no?

Porque la verdad es que eso parece prácticamente imposible. La explicación que ofrece Surowiecki es que para lograr este resultado se necesitan cuatro cosas:

- Que no haya un control central tomando decisiones.
- Que el colectivo de personas tenga mucha diversidad.
- Que las personas no se comuniquen entre ellas.
- Que haya una forma de agregar el conocimiento de todas las personas.

La falta de control centralizado es un tema recurrente en vida artificial ya que, principalmente, ofrece robustez ante fallos (si el presidente se enferma no pasa nada, porque no hay presidente).

La bondad de la diversidad también es un tema recurrente en vida artificial. Si el colectivo es homogéneo entonces los resultados son malos porque todas las personas pueden estar sesgadas en la misma dirección. Pero si el colectivo incluye personas con y sin conocimientos, con y sin experiencia y de diversas procedencias, entonces es de esperar que las opiniones particulares conformen una campana de Gauss, donde los sesgos hacia arriba se compensen con los sesgos hacia abajo, y que el resultado promedio sea muy acertado. Para ello es importante que no haya comunicación dentro del colectivo, pues de otro modo aparecerían correlaciones que impedirían la

formación de la campana de Gauss (típicamente aparecerán leyes de potencia), por ejemplo, si mucha gente decide imitar al que aparentemente más sabe, o al más charlatán.

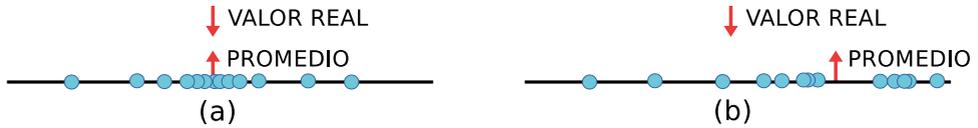


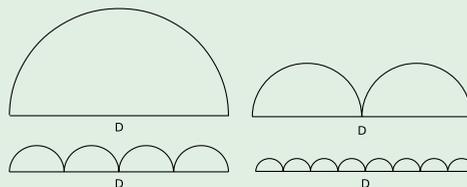
Figura 88. Dos posibilidades en el experimento de Galton

Suena bien pero, en mi opinión, Galton simplemente tuvo suerte. No hay ninguna garantía de que la campana de Gauss esté centrada en el valor correcto (como se ve en la figura 88-a). Al respecto ver también el problema 20. No hay ningún indicador que sirva para evaluar la calidad del colectivo, *a priori*. No se pueden esperar milagros de la inteligencia colectiva, y menos si se ejecutan experimentos de esta manera, suponiendo que el promedio lo va a resolver todo. Y, como más adelante reconoce el autor del mencionado libro, el experimento fracasa si se usan solo niños (Figura 88-b). O sea que la diversidad no es garantía de éxito. Hacen falta expertos. Los expertos tendrán información imperfecta y sesgos que, si no hay comunicación, tenderán a anularse. Pero todos ellos tienen conocimiento y quieren llegar a averiguar el valor real, por lo que no ocurrirá lo que vemos en la figura 88-b.

Problema 20: Aplanando la circunferencia



Dibujemos un segmento de longitud D y usémoslo como diámetro de una semicircunferencia. En su interior dibujemos dos semicircunferencias de la mitad de diámetro cada una. Repitamos infinitas veces el proceso. Al final, las semicircunferencias se confundirán con el segmento D dibujado inicialmente.



La longitud de la primera semicircunferencia es $\pi D/2$. La suma de las dos siguientes semicircunferencias es $2\pi D/4/2$. La de las 4 siguientes es $4\pi D/4/2$, y así sucesivamente. Cuando tengamos N semicircunferencias, su longitud total será $N\pi D/N/2 = \pi D/2$. Esta longitud es igual a la del segmento D , dibujado inicialmente. Por tanto $\pi D/2 = D$ de donde se deduce el verdadero valor del número π : $\pi = 2$.

¿Está usted de acuerdo?

De alguna forma, en los algoritmos evolutivos también ocurre algo así. Hay una presión selectiva empujando a que los individuos se aproximen a la mejor solución. Si se escogieran individuos al azar y se hicieran promedios no habría garantías de acercarse al óptimo. Entonces, como en toda nueva ciencia, hay confusiones iniciales. Los expertos hacen falta, pero se requiere que haya una variedad de ellos. Los ignorantes solo introducen ruido. Y por ello es que en los casos de éxito los colectivos usan sus conocimientos de una manera estructurada para lograr resultados que ninguno de sus miembros a nivel individual podría obtener. “Estructurada” significa que se usan expertos en múltiples áreas, que hay comunicación entre ellos, y que los expertos son más relevantes en la toma de decisiones que los legos. La comunicación no suele ser jerárquica como en las empresas tradicionales, sino que se permite un flujo libre de ideas y de acciones. Y en casos más complejos se requieren equipos de expertos específicamente seleccionados para atender situaciones de emergencia, como cuando en una fábrica de aviones tienen problemas completamente nuevos (Miller, 2010) o cuando quisieron rescatar a un deportista de *kitesurf* que se perdió en el Mar Rojo (Nielek, 2013) sin saber nadie su posición, sino únicamente con expertos en diversas áreas como *kitesurf*, corrientes marinas y geografía local. Aquí también muestran cómo el ruido introducido por los no expertos puede amplificarse por medio de realimentaciones positivas. Concretamente, la gente miraba las noticias en la televisión y publicaba su opinión sobre ellas en los foros de Internet. A su vez los periodistas leían los foros y daban su versión por televisión. De este modo, cualquier rumor se amplifica exponencialmente, dificultando la entrada de informaciones reales.

Pero ¿qué tipo de comunicación se requiere entre expertos que evite el sesgo de autoridad⁸¹? Esta es una pregunta fundamental, pues los agentes que conforman el colectivo tienen una inteligencia muy limitada, de modo que ella emerge gracias a la red. Pensemos de nuevo en el cerebro humano, ya que también es un colectivo: la capacidad de cómputo de una neurona es bajísima, a la vez que la arquitectura de conexiones de toda la red es tan compleja que no puede estar codificada en el ADN. De modo que se puede sospechar que hay un algoritmo parcialmente aleatorio que construye las conexiones. Su parte determinista es muy sencilla de especificar con poca información (que estará en algún lugar del ADN), y su parte aleatoria hace que cada cerebro sea distinto de otro pero aun así, todos funcionan razonablemente bien.

81 Cuando todo el mundo imita al que tiene o parece tener más conocimientos en el tema.

Averiguar cuál es ese algoritmo de generación de la red significará entender por completo el funcionamiento del cerebro. Seguramente falta poco para lograrlo, ya que después de la “década del cerebro” (1990-2000) sigue habiendo muchos otros proyectos centrados en entender el cerebro humano e incluso simularlo computacionalmente. Se puede mencionar la iniciativa BRAIN de USA, el HBP de Europa, el BlueBrain de Suiza, el Brain/MINDS de Japón, el BRAIN de China. Pero también hay iniciativas de empresas particulares y actualmente es posible acceder a APIs de simulaciones de cerebro disponibles en Internet, como el TensorFlow de Google con el que se ha derrotado al campeón mundial humano de Go; el Brain Simulator de GoodAI, corriendo sobre tarjetas gráficas con interfaz CUDA; NuPic de Numenta que simula las columnas de neuronas en el córtex; Microsoft, IBM, Yahoo, junto a muchos otros, tienen sus propios proyectos de inteligencia artificial, aunque más orientados al *Big Data* que a la simulación del cerebro, pero son caminos también muy válidos.

De todos modos lograr entender y fabricar inteligencia colectiva no es sencillo. Eric Bonabeau, un experto en el tema predijo en el 2009 que Wikipedia iba a fracasar, y hoy todavía no entiende el éxito que sigue teniendo. Y de la misma forma que Galton pensaba que el colectivo es más inteligente que cada individuo, Miller (2010) cree que “ninguno de nosotros es tan tonto como todos nosotros juntos”. Es decir, a veces la inteligencia colectiva fracasa rotundamente y pone ejemplos de ello:

- Los hormigueros la mayoría de las veces hacen un trabajo extraordinario que llamaríamos inteligente: construyen sus nidos, buscan comida y se defienden de sus enemigos, a pesar de que la inteligencia de cada hormiga es minúscula y muchas de ellas parecen empujar la comida en el sentido equivocado e incluso olvidar para donde van. Emerge un comportamiento inteligente a pesar de la estupidez de cada una de ellas. Sin embargo, a veces también emerge estupidez, como cuando hay una larga columna de hormigas caminando y dejando un rastro de feromonas. En su deambular llegan a cruzarse por algún sitio donde ya habían pasado, detectando sus propias feromonas y siguiendo el rastro. El resultado es la llamada “marcha de la muerte”, pues darán vueltas en círculos sin descanso, hasta fenecer.
- La langosta caníbal, de la que hablábamos en el capítulo “Complejidad”, es otro ejemplo de comportamiento colectivo emergente que lleva a un desastre.
- Los comportamientos colectivos humanos no están exentos de ese problema. Las burbujas económicas ocurren cuando todos los individuos quieren ganar dinero, con lo que producen un crecimiento ex-

ponencial del valor de algún bien, típicamente los bienes raíces. Y lo aceptan a sabiendas de que conducirá a una excesiva sobrevaluación del bien, hasta que llegue un momento en que nadie quiera comprarlo, cayendo en picado su precio. Y entonces se dice que la burbuja ha explotado. Todos quieren ganar, pero el comportamiento colectivo hace que todos pierdan.

Hay casos donde no es razonable dejar en manos de un reducido número de expertos la solución a un problema difícil que afecta a un colectivo mayor, y entonces se recurre a hacer votaciones democráticas, donde a cada persona le corresponde un voto. Así son los sistemas de elección de dirigentes en los países democráticos y, como muchas veces vemos, eso no significa que siempre den con una solución correcta. Además, se ha detectado que los resultados de las votaciones democráticas son predecibles con bastante exactitud cuando un grupo de expertos usa la primera técnica que vamos a ver enseguida. De modo que quizás haya que buscar una transición a un sistema democrático más sofisticado⁸². Porque hay varias técnicas para diseñar inteligencia colectiva. Las que se conocen hasta ahora son:

- **Mercados de predicción.** Cada experto hace una predicción que se transforma en una apuesta en dinero y que, al final, se convierte en un premio para el que se acerque más. Con ello se logran evitar sesgos emocionales (cuando hay dinero de por medio, nadie va a predecir que va a ganar su equipo de fútbol favorito si las probabilidades son bajas). Y quien no tiene información fiable, no apuesta nada. De modo que la información privada fiable que alguien pueda tener sale a la luz, porque se le da un incentivo económico para ello.
- **Método Delphi.** Se hace una encuesta a expertos sobre lo que piensan que va a ocurrir en un determinado tema. Y se les pide argumentos que sustenten sus opiniones. Después se publican los resultados así como los argumentos, pero de forma anónima, es decir, sin que se sepa lo que piensa cada uno de los expertos. Los expertos pueden cambiar de opinión cuando ven los resultados de la encuesta anterior así como los argumentos que apoyan cada predicción, de modo que se repite la encuesta y su publicación varias veces, hasta que se llegue a un consenso. Con ello se permite que los expertos se comuniquen,

82 Isaac Asimov anticipaba un ejemplo caricaturesco de ello en su cuento “Sufragio universal” (Asimov, 1991), que relata en un futuro distante cómo un gran computador selecciona a una única persona en todo USA que representa las características del ciudadano promedio, para que sea el único votante en las elecciones.

pero sin que medie entre ellos ninguna emoción ni ninguna presión por el respeto a la autoridad⁸³. Esto produce una realimentación positiva de las buenas ideas.

Este método es bastante intuitivo y se ha inventado varias veces. Por ejemplo, Ray Dalio (2017) lo ideó para tomar decisiones en su empresa operadora del mercado de capitales, que es la mayor de su clase en el mundo, aunque con una gran diferencia: en vez de opiniones anónimas, inculca en sus empleados la sinceridad y la transparencia, que son valores emocionalmente más difíciles de conseguir. La verdad es que es más sencillo lograr lo mismo por medio del anonimato. Lo que sí ha conseguido es ponderar las opiniones de cada funcionario en función de su credibilidad, es decir, cómo le haya ido en situaciones pasadas.

- **Votaciones.** Muchas veces se recurre a las votaciones para establecer decisiones sobre temas difíciles. Puede que entre el grupo haya expertos, pero sus votos se van a diluir entre los de la gente común, y eso va a llevar a malas decisiones. Pero hay una forma sorprendentemente fácil de separar los expertos de quienes no lo son (Vedantam, 2017), haciendo a cada persona una doble pregunta de tipo ¿qué es X? y ¿qué crees que la gente piensa que es X? Los expertos pueden diferenciar las respuestas, mientras que el común de la gente contesta lo mismo a ambas. Se mostrará con un par de ejemplos:
 - ¿Cuál es la capital de Brasil? y ¿cuál crees que piensa la mayoría de la gente que es la capital de Brasil? Si en la primera pregunta la gente contesta Brasilia y Río de Janeiro, mientras que en la segunda contesta solo Río de Janeiro, entonces la respuesta correcta es Brasilia. Y quien sabe diferenciarlos es el experto.
 - En un estanque crecen nenúfares, duplicando el área ocupada cada día. Si ocupan todo el estanque en 48 días. ¿En cuántos días ocuparon la mitad del estanque? y ¿cuántos días crees que piensa el resto de la gente que los nenúfares ocuparon la mitad del estanque? Si en la primera pregunta la gente contesta 24 y 47, mientras que en la segunda contesta solo 24, entonces la respuesta correcta es 47. Y el experto es quien sabe diferenciar las dos respuestas.
 Obviamente hay que hacer algún tipo de estadística pues es determinante que no haya nadie (o casi nadie) que responda Río de Janeiro a

83 Todos nos dejamos influir con facilidad. Los experimentos de Asch (1974) muestran personas que cambian su declaración de lo que han observado en una escena debido a que todos los demás testigos decían lo contrario (eran actores con los que se había convenido que dieran falsas informaciones).

la primera pregunta y Brasilia a la segunda. De la misma forma, cuanto más gente responde 47 a la primera pregunta que gente que responde 47 a la segunda pregunta, aumenta la probabilidad de que la respuesta correcta sea 47. Y hay más formas de distinguir sistemáticamente a los expertos, como puede verse en Pal, Harper y Konstan (2012).

Se ha diseñado *software web* de apoyo a estas metodologías, para no tener que reunir a todo el grupo de personas en el mismo sitio a la misma hora.

Y se puede considerar una forma muy general, que estoy seguro de que te va a sorprender si estás al día en las últimas tendencias en programación. Se necesitan dos fases: primero, distribuir la información entre todos los agentes que conforman la inteligencia colectiva. Y después, agrupar los resultados que cada agente ha generado. ¡Efectivamente!, es la función *collect()* de Ruby, que primero se hizo popular como el algoritmo *map-reduce* (MapReduce_Wiki, 2017) para el cálculo del *PageRank* de Google. La distribución de datos (el *map*) puede hacerse de diversas maneras, pero el objetivo es la eficiencia, por lo que la más efectiva es el árbol. Y la forma de reducir depende completamente del resultado que se desea obtener. Con este tipo de algoritmos se pueden calcular resultados de votaciones, promedios o valores de mercado que son función de la oferta y la demanda, por mencionar unos pocos. Y, por tanto, el *map-reduce* también sirve para conseguir inteligencia colectiva.

Otra pregunta válida es si podremos darnos cuenta de que existe inteligencia en el nivel del superorganismo, por encima del nuestro individual. Para entender lo difícil que es percatarse de ello, hagamos un paralelismo con una célula de nuestro cuerpo a ver si se da cuenta de que pertenece a un organismo más complejo. Terriblemente difícil, pero nosotros contamos con algo más de inteligencia que la pobre célula. Y así propongo dos indicadores de que exista una inteligencia en el nivel superior:

- Si es más fácil conseguir objetivos en grupo que solo, ello demuestra que hay sinergia en el grupo, y es un indicador débil de inteligencia en el superorganismo.
- Vamos a ir un poco más allá: si los problemas del individuo se resuelven como por arte de magia, sin necesidad de realizar ninguna acción, ello es un indicador fuerte de que el individuo se encuentra sumergido en un superorganismo inteligente, que puede predecir el futuro y anticipar las soluciones que requieran sus individuos constituyentes. Y ello ya empieza a ocurrir en nuestro mundo occidental gracias a la interacción de muchos humanos con las tecnologías de la información y las comunicaciones.

Pensemos cómo puede surgir inteligencia colectiva en una universidad. Allí hay un conjunto de profesores $\{p_1, p_2, p_3, \dots\}$, un conjunto de escuelas formadas por profesores $\{e_1, e_2, e_3, \dots\}$ y luego un enorme conjunto de comités de escuela, de programa, de planta física, de laboratorios y de investigaciones $\{c_1, c_2, c_3, \dots\}$. También se construyen comités transversales con representantes de otros comités en una jerarquía de niveles que da vértigo $\{t_1, t_2, t_3, \dots\}$. Si recordamos el primer capítulo acerca de las formas de generar complejidad, aquí vemos la más básica: combinatoria. Algunos comités generan normas para regular la acción de los profesores y de los otros comités; otros piden o emiten informes; otros solicitan u ofrecen recursos; y todo ello debe interpretarse como bucles de realimentación, con su consiguiente generación de complejidad. Como cada bucle funciona en paralelo con los otros, el resultado global es emergente, en el peor sentido de la palabra: una sorpresa continua. Porque aparecen realimentaciones inesperadas: dado que un profesor puede estar en varios comités, la información puede fluir de una manera misteriosa entre ambos o puede haber acciones de un comité que favorezcan a otro. Esto en definitiva es el comienzo de un comportamiento colectivo inteligente pero también puede dar lugar a injusticias y presiones indebidas.

La inteligencia colectiva no es fácil de lograr, y lo poco que se consigue es con mucho esfuerzo. Habitualmente los dirigentes de un grupo no saben qué hacer para cumplir objetivos, debido a la complejidad de los problemas. Suelen tener deseos, típicamente mejorar la capacidad de supervivencia del superorganismo, logrando que tenga más beneficios monetarios o expandiéndose a nuevas sucursales, pero no saben cuál es el mejor plan para llevarlos a cabo. Los dirigentes dan órdenes hacia abajo expresando sus deseos, esperando que los de abajo sepan cómo cumplirlos. Los de abajo repiten lo mismo hacia el siguiente nivel inferior. Habitualmente se acaba con operarios cumpliendo órdenes descoordinadas y posiblemente contradictorias. Pero algo harán. Las contradicciones se resuelven iterando, o sea, repitiendo una y otra vez el proceso hasta que salga algo positivo, y con ayuda de comunicación local, es decir, sin seguir el conducto jerárquico regular, sino en conversaciones de pasillo con amigos que están en otras unidades con las que se interactúa de alguna manera; por cierto, es aquí donde hay algo de inteligencia a nivel individual. Se pierden muchos recursos y tiempo en ello, pero el dirigente no se da cuenta, no le importa o incluso cree que todo ha ido bien.

A veces alguien entrevé un problema bajo una perspectiva global, e incluso propone una solución. Por ejemplo, un formato mejorado para la contratación de nuevos profesores. El formato es aceptado porque es intuitivo, fácil y rápido de usar. Pero si la secretaria que diseña los formatos no es la

misma que la que los utiliza, es decir, si quien tiene el control local no obtiene beneficios inmediatos, lo más probable es que ese formato desaparezca en la siguiente convocatoria, volviéndose al antiguo. Por ello, una estrategia para hacer nacer la inteligencia colectiva es cerrar bucles, es decir, que la persona que genere un problema sea también la que sienta los resultados de su acción. Si no hay bucles cerrados no habrá acciones correctoras.

En un hormiguero es lo mismo: las hormigas encuentran una miga de pan y algunas empujan en la dirección correcta y otras no. Ni siquiera son conscientes de esta discrepancia ni intentan resolverla. Habitualmente consiguen llevar la comida al nido, pero la ineficiencia es alta. Uno podría preguntarse por qué finalmente ganan las hormigas que empujan en la dirección correcta. Seguramente la respuesta está en la evolución, que hizo que dejaran de existir los hormigueros con los genes muy equivocados. La evolución, en ausencia de una presión selectiva fuerte, se conforma con resolver problemas aunque sea de forma ineficiente.

Pero la evolución no nos ayuda en los sistemas sociales. Sería increíblemente costoso y socialmente inaceptable, teniendo como objetivo el diseño de un formato óptimo de contratación, crear un conjunto de empresas o universidades con distintas estructuras, distintas reglas, distintos formatos, y poner una presión selectiva muy alta, por ejemplo aumentando el presupuesto de las que gradúen más estudiantes y disminuyéndolo a las que gradúen menos. Si se hace eso, muchas universidades desaparecerán, y las que queden serán las que tengan mejores formatos. Pero esta forma de generar formatos óptimos es increíblemente costosa. Además, la eficiencia de cualquier organización depende de un montón de otros procesos de modo que no hay garantías de que se logre el formato óptimo a no ser que se genere un número enorme de universidades.

¿Cómo lograr entonces inteligencia colectiva? A veces un individuo tiene una buena idea. Y además de probarla, debe convencer a los otros de que lo es y de los beneficios que traerá. No solo tiene que ser creativo sino también comunicativo. Ese tipo de cosas van mejorando el sistema, pero son muy lentas.

Las piezas que forman parte de un superorganismo deberían ser muy flexibles. O sea, lo que importa no es que la pieza pueda resolver un problema, sino que sea lo suficiente flexible como para resolver un conjunto amplio de problemas. Las neuronas son así, lo mismo que las hormigas y las personas. Un ejemplo no biológico sería las piezas de un *puzzle* versus las de un *Lego*. El *puzzle* resuelve un único problema, típicamente, una bonita foto de un paisaje. El *Lego* resuelve una multitud de problemas actuales y futuros. En este sentido, las personas en una organización son muy flexibles: pueden

realizar distintas tareas, se les puede cambiar de oficina y de funciones (las hormigas no son tan flexibles). Mientras que los sistemas de información son muy inflexibles, pues necesitan los datos en un cierto formato y orden, la *interface* con el usuario es rígida, los cálculos son siempre los mismos y el formato de salida de resultados suele ser solo uno (la mayoría de las veces es el peor formato posible, HTML, que no es interoperable programáticamente con otros sistemas de información). La ventaja del *Lego* radica en su capacidad combinatoria: las piezas se pueden conectar con otras piezas de muchas maneras. Análogamente, la tabla periódica de los elementos permite, por combinatoria, generar muchas moléculas. Si todos los átomos fueran como los gases nobles, sin capacidad de combinarse, la vida sería imposible. Por eso es que si se emplean sistemas de información inflexibles es imposible que nazca la inteligencia colectiva. Y sin inteligencia, el superorganismo en ciernes morirá, pues no podrá adaptarse a los nuevos problemas que aparezcan en su entorno.

El superorganismo necesita identificar problemas globales, pero los individuos que lo conforman solo pueden hacerlo de forma local. Es imposible para ellos identificar problemas globales. Lo único que pueden hacer desde su localidad es tratar de ver lo más lejos posible. Si en una universidad cada departamento optimiza localmente sus procesos, el resultado suele ser un montón de contradicciones e ineficiencias a nivel global. Cuando todos los individuos se dan cuenta de ello, se sentarán a negociar para ponerse de acuerdo en la manera de interoperar. Y habremos resuelto problemas locales un poco más amplios. Sin embargo, los problemas globales son más difíciles de percibir. Por ejemplo, suponiendo que la universidad prepare egresados con muy buena formación, ¿debe aumentar al máximo posible la salida de egresados? Localmente a nivel de la universidad pareciera que sí, pero ¿y a nivel social? Si estos egresados no son los que requiere la sociedad, lo único que estaremos haciendo es aumentar el problema del desempleo. Otra pregunta en la misma tónica: muchos egresados consiguen trabajo en el extranjero. ¿Es razonable que el país haya dedicado tantos recursos para formar personal que finalmente va a trabajar para otro país? A un nivel local de país, la respuesta es no. Pero a un nivel más amplio, planetario, la respuesta es sí. ¿Dónde está el nivel global? ¿Hay que situarse en el plano del universo completo? ¿Del multiverso?

No tengo respuesta, pero la pregunta se parece mucho a la que hace Dennett acerca del *yo*, como punto gordo. Dependiendo de lo amplio que seas de miras, el *yo* que quiero que sobreviva, evolutivamente hablando, puede ser mi cuerpo, mi familia, mi barrio, mi ciudad, mi país, mi planeta...

NUEVA PERSPECTIVA

Con todos los conceptos vistos hasta ahora, vamos a proponer una nueva forma de clasificar la inteligencia. Cuando hablábamos de complejidad, decíamos que todo es cuestión de grado y en el caso de la inteligencia también es así. Hay dos grados principales: inteligencia individual e inteligencia colectiva (Figura 89).

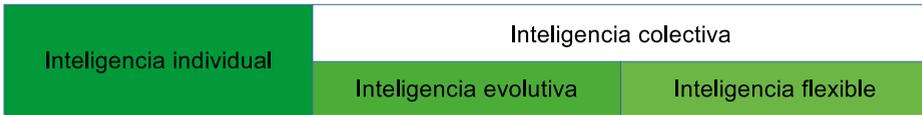


Figura 89. Inteligencia individual versus colectiva

Mucha gente no llamaría inteligente a la individual, ya que consiste en objetos únicos que no se reproducen y que tienen objetivos muy específicos. Son algo así como inteligencia solidificada. Estoy pensando en las herramientas como el destornillador, que solo sirve para apretar y soltar tornillos. El martillo para golpear clavos. Otras herramientas más abstractas que se encuadran dentro de esta categoría son el álgebra inventada por al-Jwārizmī, el algoritmo A^* inventado por Peter E. Hart, Nils J. Nilsson y Bertram Raphael en 1968 para hacer búsquedas en grafos, y el *software* de Deep Blue, especializado en jugar al ajedrez. No le pidan a Deep Blue que conduzca un automóvil por una ciudad.

Una vez que aparece la reproducción, también lo hace la inteligencia colectiva que, a su vez se puede clasificar en dos tipos: inteligencia evolutiva e inteligencia flexible (Figura 89). En la primera los individuos siguen siendo algoritmos fijos, que mueren para dar paso a otros mejores de manera que la población va aprendiendo. En este tipo, la población es más inteligente que el individuo. Es lo que ocurre con todos los algoritmos evolutivos actuales, sean biológicos o computacionales. En la segunda, los individuos se agrupan para formar superindividuos que no mueren, sino que son flexibles, se adaptan y aprenden, gracias a que llevan una población dentro de sí. Los individuos de esta población cautiva quizás si mueran en el proceso de aprendizaje, pero el superindividuo no. Es lo que ocurre con las personas y sus neuronas en el cerebro. Y la agrupación de individuos en superorganismos crea sucesivos niveles de inteligencia que no tienen por qué tener ningún techo (Figura 90).

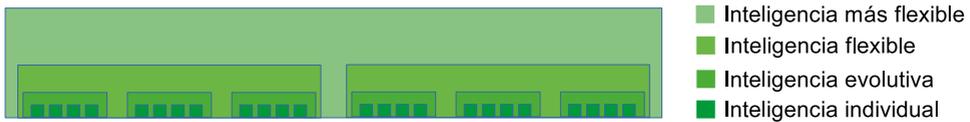


Figura 90. Emergencia de niveles de inteligencia, por agrupación de individuos en superorganismos

De hecho, el cerebro no es un lugar apacible donde simplemente transcurren los pensamientos humanos, como uno podría crear. Todo lo contrario. Las neuronas también están envueltas en una lucha terrible para sobrevivir. Seguramente por eso es que automodifican continuamente la forma de expresar su ADN (Feng, Fouse, Fan, 2007). Las que no logran nunca ninguna correlación entre sus entradas y sus salidas mueren (o modifican sus conexiones), y eso ocurre continuamente, aunque la tasa más alta es desde los 3 hasta los 16 años de vida de un niño, lo cual se puede interpretar por medio del teorema de *No-Free-Lunch* que veíamos en el libro anterior. El cerebro neuronal hace una exploración del espacio de búsqueda que le tocó, o sea, su ambiente al nacer, y descarta las neuronas y las conexiones que no le sirven para resolver problemas. Acepta que no puede resolver todos los problemas posibles y se centra únicamente en resolver de manera eficiente los problemas que encuentra en su medio ambiente. Por eso conviene darles a los niños muchos estímulos sensoriales, afectivos y cognitivos. De lo contrario, un niño aislado en un entorno pobre desarrollará un cerebro con pocas capacidades.

Estamos viendo que en el momento en que un objeto de inteligencia individual es capaz de reproducirse, pasa a formar parte de la inteligencia colectiva evolutiva. Y en el momento en que una población de inteligencia evolutiva logra aislarse del entorno y crear una representación interna del mundo, es decir, un simulador, entonces conforma una inteligencia flexible. Pero también hay un flujo en dirección contraria: las inteligencias colectivas en su quehacer pueden crear herramientas como los destornilladores y los martillos, es decir, inteligencias individuales. De hecho, un individuo de una población evolutiva es solo eso: una herramienta, una inteligencia solidificada, capaz de resolver solamente un problema. A su vez también ocurre que las inteligencias flexibles, los cerebros humanos, se agoten a ciertas edades, y pierdan la flexibilidad. Hemos escuchado personas que dicen: “lo hago así porque siempre lo hice así”. O que no pueden entender problemas nuevos. O que se empeñan en resolver nuevos problemas con herramientas viejas. O, simplemente, que les asusta el cambio. Se convierten en algoritmos fijos, como el A*. Se vuelven inteligencias individuales, inteligencias cristalizadas, solidificadas, fósiles. Herramientas.

Además, no existe un cerebro humano sino una población, la humanidad. De modo que tenemos ya otro nivel de inteligencia colectiva evolutiva. Nada impide que sigan emergiendo nuevos niveles en una especie de espiral, que genera lateralmente a su vez muchas herramientas. Y estas herramientas le sirven como apoyo inteligente para seguir creciendo. Ahora mismo estamos creando una herramienta, la inteligencia artificial que, en cuanto adquiera capacidades de autocopía, emprenderá su propio rumbo. Otras posibles consecuencias de la capacidad de autocopía en máquinas pueden verse en (CdeCiencia, 2015).

Conforme van apareciendo más niveles por agrupación de individuos del nivel anterior, el aprendizaje en cada uno se va desligando de la muerte. Los individuos de nivel superior no tienen por qué morir para aprender, aun cuando los de niveles más elementales sigan haciéndolo. En nuestras sociedades humanas parece ser que la muerte sí tiene que ver con el aprendizaje de la sociedad. Recuerdo que cuando comencé mis estudios universitarios se jubilaron muchos profesores que ofrecían asignaturas obsoletas, y dejaron el campo abierto a una modificación curricular que modernizó los estudios. En este ejemplo estoy hablando de muerte académica, que también ocurre cuando se renueva un comité, un congreso de diputados o un presidente en algún país. El superorganismo cambia de comportamiento y aprende porque sus partes constituyentes se renuevan. La democracia, tan frágil y tan asaltada en estos tiempos, tiene esa virtud. Permite la renovación, permite el aprendizaje.

RESUMEN

Siendo un concepto difuso, se empieza a aceptar que la inteligencia es la capacidad de predecir el futuro, anticiparse a él e incluso modificarlo con el objetivo de sobrevivir, lo que indica que es un producto de la evolución. Y en ese afán de anticipación del futuro, a largo plazo la inteligencia solo produce aumento de entropía, es decir, nos acerca más rápido al final del universo.

La inteligencia, al ser una cuestión de predicción, también lo es de velocidad. Cuanto más rápidamente predigas los acontecimientos futuros y cuanto más hacia adelante en el futuro te refieras, más tiempo tendrás para reaccionar por adelantado. En este sentido, las plantas son menos inteligentes que los animales. Las plantas son bastante complejas y se han descubierto fenómenos asombrosos de adaptación, comunicación e inteligencia (Gage, 2017), pero su lentitud no las permite reaccionar a problemas inmediatos. También en este sentido, los computadores son mucho más rápidos que las personas, por lo que dejo al lector que saque sus conclusiones.

La inteligencia es un producto de la evolución. Si no hay evolución, no puede haber inteligencia. Y es un resultado colectivo. Si vemos un individuo inteligente, con seguridad está formado por múltiples unidades similares (neuronas u otros sistemas). La inteligencia se desarrolla más cuanto más complejo sea el ambiente. Si el ambiente es simple, la inteligencia que allí se desarrolle también lo será, precisamente porque es un resultado de la evolución. Es decir, al tratar de sobrevivir en un cierto ambiente, hay que predecirlo y con ello surge la inteligencia. En resumen, la complejidad que se requiere para que surja la inteligencia es la misma que se necesita para que aparezca la evolución, y su nivel de sofisticación estará limitado por la complejidad del ambiente en el que se desarrolle.

El triunfo de un *software* como campeón mundial de Go, gracias a las nuevas técnicas de *Deep Learning*, ha marcado un punto de inflexión donde la inteligencia artificial supera a los humanos en tareas muy complejas y sin que sus programadores mantengan el control de qué hace o cómo lo hace.

Hemos visto las formas de razonar de los humanos, y que nuestro cerebro está orientado a resolver problemas sociales, por razones evolutivas, aunque podemos forzarlo a resolver problemas matemáticos y de lógica, pero allí muchas veces nos equivocamos. En general los humanos actuamos orientados por nuestras emociones, a las que luego buscamos una explicación lógica más o menos hipócrita. Además, para medir la inteligencia de un ente hace falta un árbitro más inteligente, de modo que ello puede dar lugar a problemas de *aliasing*.

Hemos realizado una clasificación de problemas de inteligencia artificial según su grado de dificultad, y una de las clases más interesantes son los problemas clonables. A su vez hemos visto una clasificación de tipos de mentes, y la *popperiana* es capaz de resolver los problemas clonables, gracias a que tiene dentro un simulador. Ya habíamos visto cómo funciona un simulador en un capítulo anterior. También está el tipo de mente *gregoriana*, que posee un sistema de entrada-salida universal, como por ejemplo, la mano.

Por último reflexionamos que toda inteligencia es colectiva, es decir, una agrupación de agentes independientes y diversos, que se comunican entre sí y tratan de tomar decisiones. A lo que emerge de allí, que puede ser bueno o malo, le llamamos inteligencia. Hemos mostrado ejemplos de inteligencia colectiva así como varias formas de diseñarla y un par de indicadores de su presencia.

PARA SABER MÁS

- Steven Pinker (1997). *Cómo funciona la mente*. Bogotá: Ediciones Destino.

En la primera parte explica cómo funcionan ciertos módulos cerebrales (principalmente la visión). En la segunda parte (a partir de la página 318 aproximadamente, capítulo “Buenas ideas”), explica los comportamientos humanos desde una perspectiva evolutiva, incluyendo sus bases genéticas (preprogramadas) y sus limitaciones. Un libro fascinante, para entender quiénes somos los humanos y por qué somos así. Un libro que hay que leer, si se pretende trabajar en inteligencia artificial. Libro divulgativo y fácil de leer, a pesar de su tamaño.

- Andy Clark (1999). *Estar ahí*. Barcelona: Ediciones Paidós Ibérica S. A.

Muestra el completo cambio de perspectiva que están tomando los investigadores en inteligencia artificial. Explica funciones cerebrales desde un punto de vista evolutivo. Descarta el enfoque reduccionista *top-down* habitual en ingeniería. Resuelve el grave problema de la “representación”, que nadie sabe abordarlo en inteligencia artificial clásica, convirtiéndolo en “representación no-simbólica, orientada a la acción”. Da una clara introducción al caos y a la emergencia. Habla de los nuevos robots del MIT “rápidos, baratos y fuera de control”. La cognición adulta parece lógica y proposicional pero realmente está basada en recursos corporales de tiempo real (metáforas de fuerza, acción y movimiento). Excelente libro. Abre la mente a los investigadores de la inteligencia artificial, para que exploremos nuevos caminos.

- Daniel C. Dennett (1999). *La peligrosa idea de Darwin*. Madrid: Círculo de Lectores.

Excelente libro para entender la evolución en un sentido amplio. La evolución está presente por todas partes, no solo en biología. Muchos argumentos, algunos sorprendentes. Explica incluso cómo hacerse rico, usando la evolución de una manera inusual. Libro divulgativo y fácil de leer, a pesar de su tamaño.

- Daniel C. Dennett (2000). *Tipos de mentes: hacia una comprensión de la conciencia*. Barcelona: Debate.

Empleando un enfoque evolutivo, explica los distintos tipos de mente que pueden encontrarse en el reino animal, incluyendo los humanos, y su razón de ser.

- Eric Bonabeau, Marco Dorigo y Guy Theraulaz (1999). *Swarm intelligence*. New York: Oxford University Press.

Explica en detalle los algoritmos basados en inteligencias de enjambre, incluyendo las colonias de hormigas.

- Dan Ariely (2010). *The upside of irrationality. The unexpected benefits of defying logic at work and at home*. USA: Harper Collins.

Ariely tiene varios libros donde ilustra con ejemplos la irracionalidad humana. En este comienza contándonos cosas como que cuando se pagan incentivos para alcanzar metas los resultados suelen ser mejores cuanto mayor es el incentivo. Y ello es lo racional. Pero llega un momento donde el pago es tan alto que pone nervioso a las personas, y su desempeño disminuye notoriamente. Les ocurre a los animales, a los estudiantes, a los deportistas, a los CEO y a los banqueros. La conclusión debería ser limitar los sueldos de los banqueros para lograr que sean más responsables y eviten las crisis que estamos viendo.

Después continúa hablándonos sobre el significado del trabajo. Se consiguen más objetivos cuando el trabajo se entiende y tiene algún sentido, mientras que se abandona antes cuando no tiene sentido, a pesar de que paguen igual en ambos casos. O sea que si eres un CEO que quiere desmotivar a sus empleados, la recomendación es destruir su trabajo enfrente de ellos. Muchos animales también buscan una recompensa en la comida sobre todo si ha sido difícil de conseguir, aunque los gatos son una excepción. Los *blogs* son tan populares porque esperamos que alguien nos lea. Si supiéramos que nadie nos lee, no los escribiríamos. Dividir el trabajo en parcelas especializadas hace que sea más eficiente, pero a la vez, los trabajadores pierden el sentido de lo que hacen. Lo mismo ocurre con el *software* muy particionado y manejado por muchas personas para generar un producto.

Y así sigue con una larga colección de temas donde los humanos nos mostramos en toda nuestra irracionalidad.

REFERENCIAS

Libros, artículos y enlaces web

- Ariely, D. (2008). *Predictably Irrational: The Hidden Forces That Shape Our Decisions*. USA: Harper Collins.
- Asch, S. (1974). Fuerzas de grupo en la modificación y distorsión de juicios. *Estudios básicos de la psicología social*, pp. 351-364. Barcelona: Hora.
- Asimov, I. (1950). *Yo, robot*. Barcelona: Editorial Edhasa.
- _____. (1991). *Con la Tierra nos basta*. Madrid: Martínez Roca.
- Aunger, R. (2002). *The electric meme*. New York: The Free Press.
- Bonabeau, E. (2009). Decisions 2.0: The Power of Collective Intelligence. *MITSloan Management Review*, 50(2), pp. 45-52.
- Carroll, L. (1995). *Matemática demente*. Barcelona: Tusquets Editores.
- Clarke, A. C. (1973). *Profiles of the Future: An Inquiry into the Limits of the Possible*. New York: Harper & Row.
- Csikszentmihályi, M. (2008). *El yo evolutivo*. Barcelona: Kairós.
- Dawkins, R. (1982). *The extended phenotype*. Oxford: Oxford University Press.
- _____. (1994). *El gen egoísta*. Barcelona: Salvat.
- Dennett, D. (1995). *La conciencia explicada: una teoría interdisciplinar*. Barcelona: Ediciones Paidós.
- Domingos, P. (2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. New York: Basic Books.
- Dubois, D. M. (ed). (1999). *Computing Anticipatory Systems (CASYS'99)*. USA: American Institute of Physics.
- Estructurado_Wiki (2017). Teorema del programa estructurado. Wikipedia. Recuperado el 21 de agosto de 2017. Disponible en: <https://goo.gl/yDptTw>
- Feng, J., Fouse, S. y Fan, G. (2007). Epigenetic regulation of neural gene expression and neuronal function. *Pediatric Research*, 61(5 Pt 2), pp. 58R-63R. DOI: <https://doi.org/10.1203/pdr.0b013e3180457635>
- Gage, G. (2017). Electrical experiments with plants that count and communicate. TED. Recuperado el 27 de noviembre de 2017. Disponible en: <https://goo.gl/AWUiz4>
- Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media.
- GoodAI (2015). *Brain Simulator*. Recuperado el 12 de junio de 2017. Disponible en: <https://goo.gl/Zht8GL>

- Google (2015). *Tensorflow*. Recuperado el 11 de junio de 2017. Disponible en: <https://goo.gl/M4zPhN>
- Greene, B. (2001). *El universo elegante*. Barcelona: Crítica.
- Harari, Y. N. (2016). *Homo Deus. Breve historia del mañana*. Madrid: Debate.
- Hawkins, J. (2005). *Numenta*. Recuperado el 12 de junio de 2017. Disponible en: <https://goo.gl/2S9NiK>
- _____. y Blakeslee, S. (2004). *On Intelligence. How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines*. New York: Henry Holt and Company.
- Hernández-Orallo, J. (1999). Beyond the Turing Test. *Language and Information*, 9.
- Hofstadter, D. R. (1979). *Gödel, Escher, Bach: un eterno y grácil bucle*. Barcelona: Tusquets Editores.
- _____. (1995). *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. New York: Basic Books.
- Iyyer, M., Manjunatha, V., Guha, A. et ál. (2016). *The Amazing Mysteries of the Gutter: Drawing Inferences Between Panels in Comic Book Narratives*. Recuperado el 7 de abril de 2017. Disponible en: <https://goo.gl/Qz9UQu>
- Karpathy, A., Salimans, T., Ho, J. et ál. (2017). *Evolution Strategies as a Scalable Alternative to Reinforcement Learning*. Recuperado el 7 de abril de 2017. Disponible en: <https://goo.gl/kBMJFJ>
- Kidd, C. y Palmeri, H. (2013). Rational snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability. *Cognition*, 126(1), pp. 109-114. DOI: <https://doi.org/10.1016/j.cognition.2012.08.004>
- Korb, K. B. y Nicholson, A. E. (2004). *Bayesian Artificial Intelligence*. London: CRC Press.
- Malone, T. W., Laubacher, R. y Dellarocas, C. (2010). The Collective Intelligence Genome. *MIT Sloan Management Review*, 51(3), pp. 21-31.
- Miller, P. (2010). *The Smart Swarm: How Understanding Flocks, Schools, and Colonies can Make us Better at Communicating, Decision Making, and Getting Things Done*. New Jersey: Avery Publishing Group.
- Mischel, W. y Ebbesen, E. (1970). Attention in delay of gratification. *Journal of Personality and Social Psychology*, 16(2), pp. 329-337. DOI: <https://doi.org/10.1037/h0029815>
- Nielek, R., Wawer, A. y Wierzbicki, A. (2013). Collaborative Problem Solving in Emergency Situations: Lessons Learned from a Rescue Mission. *Advanced Methods for Computational Collective Intelligence*. London: Springer. DOI: <https://doi.org/10.1007/978-3-642-34300-1>

- Neapolitan, R. E. (2004). *Learning Bayesian Networks*. New Jersey: Pearson - Prentice Hall.
- Nilsson, N. J. (1998). *Artificial Intelligence: A New Synthesis*. San Francisco: Morgan Kaufmann.
- Pal, A., Harper, F. M. y Konstan, J. A. (2012). Exploring Question Selection Bias to Identify Experts and Potential Experts in Community Question Answering. *ACM Trans. Inf. Syst*, 30(2). DOI: <http://dx.doi.org/10.1145/2180868.2180872>
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco: Morgan Kaufmann Publishers.
- Picard, R. (1998). *Los ordenadores emocionales*. Barcelona: Ariel.
- Pinker, S. (1994). *El instinto del lenguaje*. Madrid: Alianza Editorial.
- _____. (2003). *La tabla rasa*. Barcelona: Paidós.
- Punset, E. (2010). *El viaje a la felicidad*. Barcelona: Planeta.
- Quinton, A. (1999). *Hume*. Bogotá: Editorial Norma.
- Ramírez, J. C. y Marshall, J. A. R. (2015). Self-deception can evolve under appropriate costs. *Current Zoology*, 61(2), pp. 382-396.
- Rosen, R. y Klir, G. (1985). *Anticipatory Systems*. New York: Pergamon.
- Russell, S. y Norvig, P. (2014). *Artificial Intelligence: A Modern Approach*. Essex: Pearson.
- Stamos, D. N. (2008). *Evolución. Los grandes temas: sexo, raza, feminismo, religión y otras cuestiones*. Madrid: Biblioteca Buridán.
- Stephens, G. J., Silbert, L. J. y Hasson, U. (2010). Speaker–listener Neural Coupling Underlies Successful Communication. *PNAS*, 107(32), pp. 14425–14430. Recuperado el 7 de abril de 2017. Disponible en: www.pnas.org/cgi/doi/10.1073/pnas.1008662107
- Suárez, É., Bucheli, V. A., Zamara, R. y García, A. (2015). Collective Intelligence: Analysis and Modeling. *Kybernetes*, 44(6/7), pp. 1122-1133, DOI: <https://doi.org/10.1108/K-11-2014-0245>
- Surowiecki, J. (2004). *The Wisdom of Crowds*. USA: Doubleday.
- Trivers, R. (2011). *The Folly of Fools: The Logic of Deceit and Self-deception in Human Life*. New York: Basic Books.
- Whitby B. (1996). *Reflections on Artificial Intelligence*. Exeter, UK: Intellect Books.
- MapReduce_Wiki (2017). *MapReduce*. Recuperado el 16 de julio de 2017. Disponible en: <https://goo.gl/MUjzMD>

Películas y videos

- Cameron, J. (1984). *The Terminator*. USA: Hemdale Film. Cinema 84, Euro Film Funding, Pacific Western.
- CdeCiencia (2015). *¿Qué es la vida?* Recuperado el 8 de octubre de 2017. Disponible en: <https://goo.gl/MFs7yp>
- Dalio, R. (2017). *How to build a company where the best ideas win*. TED. Recuperado el 7 de septiembre de 2017. Disponible en: <https://goo.gl/WoP2rM>
- Garland, A. (2015). *Ex Machina*. Reino Unido: DNA Films, Film4.
- Vedantam, S. (2017). Ep.64: I'm right, you're wrong. *Hidden Brain*. Podcast. Recuperado el 3 de octubre de 2017. Disponible en: <https://goo.gl/tHMPn6>

Tesis y trabajos de grado en EVALAB

- Narváez, S. (2016). *Las máquinas y el lenguaje: desarrollo e implementación de un sistema para la asociación entre el lenguaje natural y eventos en un mundo virtual utilizando NUPIC*. Cali: Universidad del Valle.

CONSCIENCIA

Cuando su padre le comunicó su alarma por haber olvidado hasta los hechos más impresionantes de su niñez, Aureliano le explicó su método, y José Arcadio Buendía lo puso en práctica en toda la casa y más tarde lo impuso a todo el pueblo. Con un hisopo entintado marcó cada cosa con su nombre: mesa, silla, reloj, puerta, pared, cama, cacerola. Fue al corral y marcó los animales y las plantas: vaca, chivo, puerco, gallina, yuca, malanga, guineo. Poco a poco, estudiando las infinitas posibilidades del olvido, se dio cuenta de que podía llegar un día en que se reconocieran las cosas por sus inscripciones, pero no se recordara su utilidad. Entonces fue más explícito. El letrero que colgó en la cerviz de la vaca era una muestra ejemplar de la forma en que los habitantes de Macondo estaban dispuestos a luchar contra el olvido: Esta es la vaca, hay que ordeñarla todas las mañanas para que produzca leche y a la leche hay que hervirla para mezclarla con el café y hacer café con leche. Así continuaron viviendo en una realidad escurridiza, momentáneamente capturada por las palabras, pero que había de fugarse sin remedio cuando olvidaran los valores de la letra escrita,

GABRIEL GARCÍA MÁRQUEZ (CIEN AÑOS DE SOLEDAD).

What is mind? No matter. What is matter? Never mind,

GEORGE BERKELEY.

Si no lo hubiera creído, no lo habría visto,

MARSHALL McLuhan.

Y seguimos subiendo por la escalera de la complejidad (Figura 91). Acabamos de llegar a un nuevo piso estable, generado por la emergencia de un fenómeno extraño y misterioso: la consciencia. Vamos a dar una breve in-

troducción a lo que se sabe acerca de ella, a algunos experimentos e hipótesis relevantes, pero no daremos una revisión histórica y quien esté interesado puede remitirse a libros de filosofía bastante agradables de leer como los de Daniel Dennett (1991, 1995, 1996 y 2003). La consciencia es algo tan subjetivo y misterioso que durante mucho tiempo la ciencia se negó a considerarlo como un tema de estudio. Pero si, como hemos visto, la inteligencia es algo algorítmica, ¿no será que la consciencia es otro algoritmo? Enseguida haremos una propuesta en este sentido que, hasta donde yo sé, es completamente nueva y se diferencia radicalmente de las demás, pues muestra la consciencia como un fenómeno emergente bajo una interpretación computacional.

Comencemos pues, por los hechos históricos más relevantes.

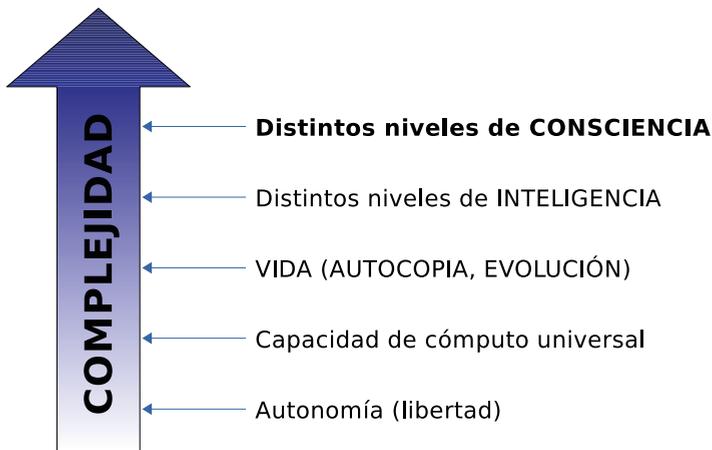


Figura 91. No se sabe cuántos bits se requieren para lograr consciencia, pero no son muchos, pues solo se precisa un bucle de realimentación

Se nos ha presentado siempre la consciencia como algo misterioso que está más allá del entendimiento humano. En cierto modo, así es. Se trata de una de tantas limitaciones *gödelianas* que tenemos. Recordemos que cuando Gödel logra codificar la aritmética de números enteros para que hable de sí misma, aparecen proposiciones que son verdaderas pero indemostrables. De manera análoga, cuando enfocamos nuestra inteligencia para que investigue sobre ella misma, lo logramos hacer, pero aparecen fenómenos nuevos que escapan a nuestra comprensión. La consciencia es real, está allí y la sentimos trabajar y, sin embargo, no entendemos qué es. De forma similar a las proposiciones de Gödel, para lograr comprenderla hay que dar un salto por fuera del sistema, que es lo que se presentará en este capítulo.

Ha habido escasos avances en la comprensión de la consciencia hasta hace relativamente poco. El dualismo cuerpo-mente de Descartes con sus

connotaciones religiosas ya está superado. Pero tuvo su importancia pues significó reconocer dos componentes, el *hardware* y el *software*, como diríamos hoy. El *hardware* se sabía que estaba formado por muchos módulos, como el corazón, el hígado o el páncreas. Pero se suponía que el *software* era monolítico, y solo muy recientemente se ha ido entendiendo que también consta de módulos. Actualmente, gracias a la tomografía cerebral en sus diversas variantes se han obtenido mapas del cerebro con las funciones específicas de cada parte así como correlaciones entre ellas y lo que decimos que estamos haciendo o sintiendo. Esto es lo que se llama la interpretación correlacional de la consciencia.

Incluso antes de disponer de estas herramientas electrónicas, las enfermedades y accidentes que dañaban una parte del cerebro permitían a los investigadores hacer deducciones basándose en las funcionalidades que se iban perdiendo.

Uno de los casos más emblemáticos fue el del obrero de ferrocarril Phineas Gage al que una varilla de hierro le atravesó la cabeza en un accidente en 1848. En el hospital le salvaron la vida, pero sus amigos decían que desde entonces ya no era el mismo. Se había convertido en una persona grosera y sin consideraciones para nadie. De allí y otros casos similares se dedujo que los dos lóbulos frontales, que fueron la región cerebral que sufrió todo el daño, eran los encargados de las funciones de alto nivel de la personalidad, como el autocontrol. Con ello se derrumba el mito de la unicidad de la consciencia y comienza a entenderse que el cerebro contiene módulos con funciones bien definidas y que el resultado de la operación conjunta de todos ellos es lo que produce inteligencia y consciencia. A partir de allí quedaron abiertas las puertas para que la ciencia investigase cuáles son esos módulos, cuál es su funcionalidad y cómo interaccionan entre sí.

Con estos hallazgos se afianza la psicología evolutiva cerrando un capítulo oscuro de un área de conocimiento, la psicología tradicional, que hasta ese entonces no llegaba a ser una ciencia. La psicología evolutiva sustenta que los módulos cerebrales aparecieron de forma evolutiva para que los genes pudieran solucionar problemas de supervivencia y reproducción en un entorno cada vez más complejo. Los genes fabrican cuerpos con cerebros que vienen ya preprogramados con algunos comportamientos simples de tipo “si se acerca un individuo muy grande y con la boca abierta, huye”, a los que llamamos instintos. Pero no pueden lidiar con comportamientos complejos en los que haya que analizar las sutilezas del entorno y tomar decisiones que dependen de múltiples condiciones. No hay genes que codifiquen “si se acerca una persona, y la persona es tu jefe, y está de mal humor, y hay rumores de despidos en tu empresa, entonces dile sí a todo”.

Uno de los pilares de la psicología evolutiva, como nos explica Steven Pinker en su libro *Como a mente funciona* es que cada módulo cerebral evolucionó para lidiar con estas sutilezas (si esa persona es o no tu jefe, cómo evaluar qué grado de humor tiene hoy o si los rumores que corren se pueden considerar fiables). Y, dado que los cambios culturales humanos son muy rápidos mientras que la evolución genética es lenta, es razonable suponer que los genes estén solucionando grandes problemas de hace miles de años, pero no sean capaces de lidiar correctamente con los de hoy en día. Por eso nos asustan instintivamente las culebras mientras que no le tenemos ningún miedo a los automóviles, a pesar de que las primeras ya prácticamente no causan víctimas entre la población humana mientras que los segundos son una de las principales causas de muerte en cualquier parte del mundo. Hay genes para reconocer y sentir miedo de serpientes y arañas, pero no ha habido tiempo de que se desarrollen genes que hagan lo mismo con los automóviles. Ni lo llegará a haber, porque cuando pasen los miles de años que requiere la evolución, para entonces las culturas humanas habrán inventado otros medios de locomoción irreconocibles en nuestro presente.

Lo anterior sucede también con muchas situaciones de la vida cotidiana. Los celos tuvieron razón de ser en los humanos (y siguen teniéndola en los animales) en una época donde nadie quería cargar con la alimentación y cuidado de unos hijos que potencialmente podrían ser de otro macho. Hoy día los análisis de ADN pueden resolver fácilmente esa cuestión sin necesidad de violencias, pero todavía sentimos celos porque esos genes siguen estando allí. Y se puede decir lo mismo de la mayoría de sentimientos negativos como la ira, la envidia o el impulso de amedrentar a otros. Los tribunales de justicia, la propia psicología evolutiva o la teoría de juegos, que son herramientas culturales humanas, vuelven inútiles estos sentimientos primitivos. Pero los genes no lo saben y eso crea conflictos mentales en las personas que no entienden que sus sentimientos están preprogramados para resolver problemas que ya no existen. Debería ser de cultura general tener unas bases mínimas en psicología evolutiva desde el colegio, para evitar mucha violencia y sufrimiento en la vida de adultos. Y debería ser obligatorio pasar exámenes al respecto si se aspira a representar un colectivo en algún tipo de organización económica, política o social, pues vemos en las noticias cómo muchos grandes dirigentes del mundo fallan estrepitosamente, comportándose como niños (o como simios) en discusiones que podrían resolverse de forma más civilizada.

Para ejemplificar voy a remitirme a un conocido presidente de un gran país, elegido hace pocos días. Resulta que este presidente toma decisiones de manera visceral, dictadas por su carácter, es decir, por sus genes. No

analiza primero pros y contras, no cuenta con la información adecuada de la situación actual y de las reglas del juego, y además su principal estrategia es repartir amenazas e insultos para todos, tanto a los críticos internos como a los externos e incluso a los amigos. Lo más probable es que no dure mucho así. En teoría de juegos se sabe que las amenazas tienen sentido si se posee la capacidad y la voluntad de cumplirlas. O, en caso contrario, si el oponente no sabe que te falta esa capacidad o voluntad. Pero las decisiones económicas proteccionistas de Trump (vaya, se me escapó) y sus correspondientes amenazas a diversos países no tienen ningún asidero, pues de llevarse a cabo el principal perdedor será su propio país. Nadie creará esas amenazas. Este no pasa de ser un ejemplo memorable, pero no es la excepción sino la culminación, pues muchos países tienen dirigentes con características similares. Es hora de decir ¡basta! Estamos ya en el siglo XXI, en la era de la información. Los conocimientos sobre la teoría de juegos y la psicología evolutiva están disponibles para todos los públicos, por lo que no deberíamos elegir a gente con tanta ignorancia, que incluso se vanaglorian de no leer ningún libro. Educar en estas áreas desde la niñez es muy importante, por el bien de la humanidad.

Volvamos al relato de casos espectaculares que revelan algún tipo de malfuncionamiento en los módulos cerebrales. La prosopagnosia es la incapacidad de reconocer rostros. Es una enfermedad relativamente común y sin mayores consecuencias, pero hay algo similar y mucho más inquietante que es lo que se llama el síndrome de Capgras, que se produce como consecuencia de la demencia senil o de golpes fuertes en la cabeza. El sujeto que lo padece cree que las personas de la familia han sido sustituidas por impostores, que son idénticos en apariencia. Lo que ocurre realmente es que la capacidad consciente de reconocimiento de rostros sigue intacta (no hay prosopagnosia), pero dejó de funcionar la capacidad de reconocimiento afectiva, que es inconsciente y que posiblemente esté mediada por los olores. En el libro *El hombre que confundió a su mujer con un sombrero* del neurólogo Oliver Sacks (2002) se comentan en detalle varios casos similares.

Otro caso que merece la pena comentar por lo sutil, es el del falso ciego (el término científico es anosognosia visual). El sujeto que lo padece alega que no puede ver. Según el relato de Dennett (1995, p. 371), lo que hizo sospechar a los médicos que ello podría ser un engaño fue que cuando al paciente se le ponía a caminar en una gran habitación vacía que solo tenía una silla, invariablemente se dirigía hacia ella, chocándose, y así demostrando (en su propio modo de entender la situación) que él estaba ciego. Obviamente, si se dirige siempre hacia la silla es porque sabe dónde está. La puede ver. El caso se resolvió cuando se dieron cuenta de que tenía dañada

la consciencia de la visión, no la visión en sí. De alguna manera podía ver a nivel inconsciente, pero no era consciente de ello. Y algo parecido nos pasa a todos, pues en cada ojo hay un punto bastante cerca del centro de la visión, que no tiene fotorreceptores debido a que es por allí desde donde sale el nervio óptico hacia el cerebro. Lo deberíamos ver como dos círculos negros en nuestro campo de visión, pero lo cierto es que nadie se da cuenta porque allí están suprimidos los circuitos neuronales que nos harían conscientes de ello. Como dice Dennett, “somos ciegos a la ceguera del punto ciego”.

Y el último ejemplo, quizás el más sorprendente de todos porque pone en duda quiénes somos, si somos libres y para qué sirve realmente la consciencia, también descrito por Dennett en el mismo libro (p. 177): a una persona se le han colocado electrodos en su cerebro, con su consentimiento, que miden la actividad motora de la mano derecha. Se le pide que pulse un botón de un proyector de diapositivas cuando en un reloj la aguja pase por las doce. Pero él no sabe que el botón del proyector realmente no hace nada, y que las diapositivas cambian cuando el electrodo detecta el comienzo de la actividad neuronal que va a contraer sus músculos de la mano para pulsar el botón. La persona se sorprende de que la diapositiva cambie justo antes de que ella tome la decisión de pulsar el botón. Le parece magia. El retraso medido desde que comienza la actividad neuronal hasta que la persona es consciente de que va a realizar la acción está entre 300 ms y 500 ms, es decir, ¡medio segundo! Pero entonces ¡la consciencia no es quien toma las decisiones!

Sin necesidad de hacer montajes tan invasivos, todos hemos experimentando este efecto cuando mantenemos una conversación informal en un ambiente agradable entre amigos. Uno habla y habla espontáneamente hasta que de repente dice algo inconveniente, o que se suponía era un secreto. Justo cuando tus palabras están saliendo de tu boca, te das cuenta de que has metido la pata. Esa es la consciencia, que no está a cargo de la conversación, sino solo supervisándola, y que da la orden de callarse demasiado tarde.

Estos experimentos y muchos otros ejemplos involucrando otras partes del cerebro están bien documentados en artículos científicos y en libros de divulgación como los de Damasio (2006), Dennett (1995) y Sacks (2002). Ello nos conduce a pensar que los módulos cerebrales evolucionaron para resolver problemas específicos que nuestros ancestros enfrentaron hace miles de años. Y que el cerebro no está diseñado para el arte o las matemáticas, de modo que si hemos logrado entenderlas es porque se trata de un efecto secundario producido por tener tantos módulos diversos con tanta capacidad de cómputo. Pero, además, ante ciertas situaciones no habituales para nuestros ancestros, los módulos pueden fallar estrepitosamente. Ejemplos divertidos de ello son las ilusiones ópticas (Díaz, 2015) y acústicas (Mikrojo,

2016a, 2016b, 2017a, 2017b), y también los errores en la forma de razonar de los humanos (Pinker, 1997, p. 357), como vimos en el capítulo anterior.

Entonces la mente está formada por muchos módulos, que pueden funcionar bien o mal. Uno de esos módulos es la consciencia y si reflexionamos nos damos cuenta que no siempre está presente. Cuando dormimos o estamos bajo anestesia, no somos conscientes. Y, cuando está presente, lo hace en diversos grados, por ejemplo, cuando vamos conduciendo un automóvil en dirección a un teatro a la vez que estamos sumergidos en una conversación con un acompañante, es posible que llegemos a nuestra casa o trabajo, en vez de al teatro, para nuestra propia sorpresa. Nuestro cuerpo y nuestra mente inconsciente tomaron las riendas del automóvil y lo llevaron al “lugar de siempre”, sin que la consciencia interviniera para nada.

Por cierto que aunque un trozo de mi cerebro sea consciente y otro no, ambos forman parte integral de mí mismo. No hay que pensar en el inconsciente como algo o alguien externo, ajeno o incluso traidor a mí. Es también una parte de mí, pero de la que sé poco porque no puedo interrogarla. Una bonita conjetura acerca de la consciencia, alabada y criticada por igual, propuesta por Julián Jaynes y explicada en Dennett (1995), supone que ella es un fenómeno relativamente reciente entre los humanos: tiene entre 500 y 3000 años de antigüedad, lo cual es muy poco evolutivamente hablando, y podría decirse que estamos todavía en el proceso de su emergencia.

Y ello implica, aunque es políticamente incorrecto decirlo, que haya humanos más conscientes que otros, por motivos puramente genéticos, de la misma manera que los hay más inteligentes que otros por los mismos motivos. La hipótesis de Jaynes, conocida como “de la ruptura de la mente bicameral”, señala que los dos cerebros que tenemos hace miles de años estaban bastante más desunidos que ahora, y solo había un flujo efectivo de información del lado derecho al izquierdo. Recordemos que los módulos del habla se encuentran principalmente en el cerebro izquierdo de las personas diestras. Además, cada lado del cerebro recibe información sensorial y controla los músculos del lado contrario del cuerpo.

Uno de los indicios a favor de esta hipótesis es que la palabra “consciencia” con su significado moderno solo existe desde hace unos 500 años. Por supuesto, uno puede aprender a usar unos palillos chinos en China sin saber cuál es su nombre. Pero se hace muy difícil de sostener lo mismo para la consciencia, puesto que ser consciente implica la capacidad de reflexión sobre uno mismo y sobre su consciencia. Tener una palabra para ello es ineludible. El neurólogo Mariano Sigman (2016) ha realizado un análisis automático del uso de la palabra consciencia y de otras palabras relacionadas, en textos desde la antigua Grecia hasta hoy, corroborando esta hipótesis.



Fuente: Rebecca Goldstein (2005). Disponible en: <https://goo.gl/SERxVN>

Personaje 8

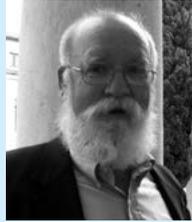
Steven Pinker (1954-)

Steven Pinker es un lingüista y psicólogo evolutivo canadiense. De hecho, sus dotes como escritor divulgador de ciencia han logrado llevar al público la idea de que la evolución es la única vía para darle sustento científico a la desprestigiada psicología tradicional. Según Pinker, la mente tiene módulos funcionales que han evolucionado para responder a las necesidades de los humanos y, entre ellos, destaca el módulo del lenguaje, tal y como decía Chomsky. Con ello se opone frontalmente a los psicólogos tradicionales que consideran al cerebro de un niño como una tabla rasa, en la que se puede construir cualquier cosa por medio de un aprendizaje adecuado. Por el contrario, Pinker piensa que el cerebro de un niño al nacer viene dotado de una serie de capacidades innatas que pueden ser aprovechadas o no, por medio de aprendizaje, pero que, debido a ello hay cosas que puede aprender (como cualquier lenguaje humano, cosa que el cerebro de los demás primates no puede), hay cosas que no puede aprender, y hay cosas que puede aprender con dificultad (como las matemáticas avanzadas), en función de los módulos innatos que posea. Otra conclusión interesante es que los 10 000 años que llevamos desde que existe la agricultura y los 200 desde que existe la industria no son suficientes para cambiar los genes, de modo que nuestro carácter y nuestra forma de pensar es adecuada para los problemas que teníamos en el Pleistoceno. Otros científicos que trabajan o apoyan la psicología evolutiva son David Buss, Leda Cosmides, Martin Daly, Margo Wilson, Daniel Dennett, Richard Dawkins y el divulgador científico Roberth Wright, aunque también hay críticos que piensan que debe reformularse con más rigor, como Buller (2005).

También son muy interesantes sus trabajos en lingüística, donde consigue armar un árbol evolutivo de todas las lenguas humanas, que coincide muy bien con las migraciones, y hacer arqueología de palabras. Una de las ideas principales es que las sentencias de cualquier lenguaje (incluyendo los de computador) son árboles, pero hay que serializarlos para convertirlos en habla. El proceso de conversión depende de cada lenguaje y allí identifica unos 20 parámetros que controlan cómo hacerlo (si el verbo va al final o separando sujeto y complemento; o si el adjetivo va antes o después del nombre). Los niños, al aprender la lengua de sus padres, además de identificar el vocabulario lo que hacen es ajustar esos 20 parámetros.

Otro de los indicios son los escritos de la Grecia clásica. En la *Iliada* y la *Odisea*, Homero dice que sus personajes escuchaban las voces de las musas y de los dioses. Si vivieran hoy, serían diagnosticados de esquizofrenia. Pero en aquella época podría significar que el hemisferio derecho estaba comunicándose con el izquierdo, que es el que tiene capacidad de hablar al mundo exterior. El izquierdo es el “*front-end*”, como se dice en computación,

y explica a las demás personas que está oyendo voces, aun cuando no sabe de dónde, porque no tiene un bucle hacia atrás que le permita dialogar y conocer nada acerca de su lado derecho.



Fuente: CC BY 2.0, David Orban (2006). Disponible en: <https://goo.gl/NDKNtR>

Personaje 9

Daniel Dennett (1942-)

Daniel Clement Dennett es un filósofo extraordinario y un gran divulgador. La filosofía, que al principio fue el motor de las ciencias, con el tiempo se ha convertido en un enredo de palabras sin sentido. Hay unos pocos filósofos que se apartan de ese esquema y entre ellos sobresale Dennett, que siempre trabaja de la mano con científicos, neurólogos e ingenieros. Es decir, hace filosofía pero partiendo de datos experimentales. Por el contrario, se dice que los filósofos tradicionales tienen los pies firmemente asentados sobre el aire.

Entre los libros de Dennett destaca *La peligrosa idea de Darwin*, donde explica la evolución y pone múltiples ejemplos, no solo en el ámbito biológico. Todavía hay un cierto debate entre legos sobre si la evolución es verdad o no. A ellos les recomiendo que aprendan a programar y escriban un algoritmo evolutivo, con lo que se les despejarán las dudas para siempre de lo bien que funciona. Otros libros interesantes son *Tipos de mentes*, *La consciencia explicada* y *La libertad evoluciona* donde elimina muchos mitos sobre estos temas, dejándonos la idea de que el yo, los qualia y la libertad, son solo una ilusión. Un libro anterior, *The mind's I*, escrito con Douglas Hofstadter, explora muchos tópicos sobre la consciencia de una manera inusual: primero incluyen un texto de un reconocido autor como Borges, Turing, Dawkins, Smullyan, Searle y otros, y luego lo analizan o critican, argumentando para alcanzar su objetivo final, que es demostrar que la mente se puede mecanizar, es decir, se puede lograr la inteligencia artificial en sentido fuerte. Recuerdo especialmente los cuentos de ciencia ficción de Stanislaw Lem, terriblemente divertidos y anticipadores.

Daniel Dennett amplía esta hipótesis respecto a la capacidad del habla humana: un protohomínido grita para atraer la atención de otros de su misma especie, quizás porque encontró comida y necesita ayuda, quizás para avisar de algún peligro. En alguna ocasión, está solo y no hay nadie que responda a sus gritos. Pero de todos modos son oídos por su oreja izquierda, conectada a su cerebro derecho que puede reaccionar, tratando de ayudarlo. Este puente de comunicación externo izquierdo → derecho, con el tiempo podría irse interiorizando, ya que tendría obvias ventajas: conseguir que la comunicación fuera privada. De modo que podría haber presión evolutiva empujando en esa dirección (aun hoy día todos conocemos de personas que

necesitan decirse las cosas en voz alta para lograr ordenar sus pensamientos). Finalmente, con la comunicación bidireccional interna terminada, aparece un nuevo bucle que, cómo no, va a producir la emergencia de nuevas e interesantes propiedades: la consciencia.

Porque la consciencia es muy rica y tiene muchas facetas y grados de profundidad:

- **Consciencia del entorno.** Puedo usar los sentidos para percibir lo que me rodea, y detectar situaciones beneficiosas y perjudiciales, para acercarme a las primeras y alejarme de las segundas. Para ello hay que tener una memoria donde almacenar y clasificar situaciones anteriores similares. Además está la integración sensorial: cuando oímos una campana y vemos una campana, sabemos que se trata del mismo objeto. Y relacionado con ello está la capacidad de atención, es decir, poder centrar la consciencia en un único objeto en cada momento, ignorando todo lo demás. La mayoría de animales tienen este tipo de consciencia.
- **El concepto del yo.** Hay algo misterioso en ello, pero la idea principal es que puedo identificarme a mí mismo, separándome del entorno. Aparentemente, el *yo* es quien gobierna las acciones del cuerpo, ejerciendo su libertad (enseguida veremos que eso no es realmente así). Muchos filósofos distinguen varios matices en este concepto: el *yo*, el *ego*, el *mí mismo*⁸⁴... Nosotros los englobaremos en una sola cosa. Los animales mamíferos suelen tener una idea del *yo*. Se ha comprobado que si se le da un espejo a un orangután (que no ha visto tal artefacto en su vida), en poco tiempo lo usa para observarse a sí mismo, hacer muecas, ayudarse de ello para rascarse la espalda y cosas así. Ello significa que se reconoce en su reflejo, por lo que tiene un concepto de él mismo. Otros animales no son así, y se asustan al verse reflejados, pues creen que se trata de un rival.
- **Autoconsciencia.** Soy capaz de pensar sobre mí mismo y hacer introspección y análisis de lo que pienso y siento, de lo que hice o pude hacer o haré.
- **Los *qualia***⁸⁵. tengo una representación interna de mis sensaciones que es subjetiva, que no puede compararse con las de otros seres similares. Ello es algo bastante misterioso. Siempre se pone como ejemplo “la sensación que tenemos del color rojo”. No la podemos explicar. Es completamente nuestra. Y no sabemos, ni podremos saber de ninguna

84 Y de donde deriva una palabra tan interesante como “ensimismarse”.

85 *Qualia* es una palabra que viene del latín, está en plural, y su singular es *quale*.

manera, si cuando dos personas miran el mismo objeto rojo experimentan la misma sensación en su mente. Lo mismo puede decirse de la sensación de hambre, la sensación de llegar tarde, la sensación de estar cansado o feliz y tantas otras.

- **La subjetividad de la consciencia.** Está muy ligada a lo dicho sobre los *qualia*, pero es más general. Es la sensación misteriosa de estar experimentando un *yo* sumergido en un mundo de sensaciones que, aunque están relacionadas con parámetros que pueden medirse (si estoy en la calle a -20 grados experimento frío), la relación de mí con el mundo es completamente personal, indescriptible e incomparable con la de otras personas. Podemos sentir. No es que estemos viendo una película, es que estamos dentro de la película experimentando un flujo continuo de sensaciones. Es lo que el filósofo David Chalmers (1995) llama el problema principal de la consciencia⁸⁶.

La primera de estas características es la más abordable desde un punto de vista computacional y ya hay muchos trabajos que apuntan a resolverla, principalmente en el área de visión computacional. Los problemas que involucra son meramente técnicos. Pero las otras tres características tienen un halo de misterio que impide un acercamiento ingenieril. Sin embargo, aquí presentaré unas ideas que podrían cambiar esa situación.

Durante mucho tiempo se pensó que solo los humanos tenían consciencia, pero en los últimos años se ha revisado esta idea. Los animales tienen consciencia en mayor o menor grado. Recomiendo volver a ver la excelente presentación de Frans de Waal en TED (2011) donde explica que los fundamentos de la moralidad humana ya se encuentran en muchos primates y que la causa es evolutiva. Los chimpancés pueden cooperar y exigir cooperación a otros, así como los elefantes, que vuelven periódicamente a los sitios donde hubiera muerto algún miembro querido de su familia. Los delfines jóvenes se agrupan para molestar y golpear a algún otro; hacen *bullying*, como los adolescentes humanos. Y quienes tienen algún animal doméstico saben perfectamente que sienten, sufren, se alegran igual que su dueño. Los perros en algunos aspectos son muy poco inteligentes, pero en otros pueden demostrar comprensión y voluntad para realizar acciones que no son mecánicas sino que identificamos en ellas que “hay alguien allí dentro, decidiendo”. Ya se han hecho escáneres cerebrales de perros donde se descubre que, cuando están alegres, están pensando en sus amos. Algo similar ocurre con gatos y algunos pájaros como los cuervos, de lo que te das cuenta si eres buen ob-

⁸⁶ *The hard problem of consciousness.*

servador. La consciencia en sus múltiples facetas es una cuestión de grado, y no de todo o nada, un tema recurrente a lo largo de este libro.

Ha habido varios intentos para tratar de explicar de dónde surge el *yo*. Recuerdo muy bien el primero que llegó a mis oídos: se había descubierto una oscilación en el cerebro, en forma de un barrido de disparos neuronales en el córtex, de una frecuencia en torno a los 40 Hz. El famoso biólogo co-descubridor de la estructura del ADN, Francis Crick, proponía en 1990 que ese barrido de 40 Hz era el que producía la integración sensorial y daba lugar a la experiencia unificada subjetiva de la consciencia.

Eso da que pensar. ¿40 Hz es un número mágico y si fuera 258 Hz ya no funcionaría? ¿Si diseño un circuito electrónico que oscile a esa frecuencia tendrá consciencia? ¿O será que hay una reacción química clave en el cerebro que es la que produce la consciencia? La respuesta es obviamente negativa a todo ello. Tiene mucha razón Chalmers en sus preguntas (no tanto en sus respuestas) acerca de que no existe ningún fenómeno físico o químico que tenga la potencia causal para argumentar que genera consciencia. Estamos hablando de niveles distintos. No hay forma de argumentar que un barrido eléctrico sincronizado o una reacción química puedan producir algo tan extraño como la consciencia. Y eso me llevó a pensar: ¿qué tipo de razonamiento sería el adecuado, convincente, y tendrá la suficiente capacidad explicativa para lograrlo? Tiene que ser algo bastante distinto, no físico ni químico, quizás matemático, quizás sistémico como los bucles de realimentación, quizás computacional.

Por cierto que los 40 Hz lo que más bien indican es que el cerebro es digital a un cierto nivel. Obviamente, en el bajo nivel los potenciales de acción de cada neurona cambian de manera continua, como cualquier voltaje, pero la integración de la información sensorial y su percepción a alto nivel es discreta. Cada 25 milisegundos (la inversa de 40 Hz) el cerebro muestrea sus sensores y toma decisiones. Por ello, los sonidos cuya frecuencia es mayor a 20 Hz se perciben correctamente, mientras que en los de frecuencia menor se percibe cada onda como un evento individual, como un golpeteo. Y también por ello las películas de cine deben ir a 24 fotogramas por segundo (o más rápido) pues para cadencias más lentas comienzan a notarse las fotos de forma aislada y se pierde la sensación de continuidad. Aunque estos números no coinciden exactamente, si lo hacen de forma aproximada. Hay que tener en cuenta que la frecuencia de 40 Hz no es fija y puede llegar hasta 70 Hz, variando según nuestro estado de atención y también entre una persona y otra. Además, hay varios circuitos cerebrales involucrados y cada uno puede tener una frecuencia distinta. Laura Spinney (2015) nos ofrece una explicación más detallada de todo ello.

UNA PROPUESTA COMPUTACIONAL PARA ENTENDER LA CONSCIENCIA

En este estado de cosas podemos usar los conceptos de la vida artificial como vía para proponer una respuesta. Concretamente podemos imaginar qué le pasaría a un futuro robot que hubiera sido fabricado como resultado de un proceso evolutivo, en el sentido de que tiene motivaciones cableadas en *hardware* para sobrevivir, buscar situaciones beneficiosas y huir de los peligros. Vamos a imaginar que ya tiene inteligencia suficiente para identificar objetos en su campo visual, memorizar situaciones y hacer predicciones sobre el futuro. Todo esto son cosas técnicas que hoy día ya se están logrando y no hay allí ninguna magia ni misterio. Por ejemplo Google ya ha logrado un *software* de *deep learning* capaz de identificar gatos en videos, sin ayuda externa ni entrenamiento supervisado (Quoc *et ál.*, 2012).

La consciencia se ha entendido desde hace tiempo que debe basarse en algún tipo de bucle, como por ejemplo Edelman (1998) con sus mapas neuronales reentrantes y Hofstadter (2007) con sus bucles autorreferenciales. Ya adelanto que esta nueva propuesta se basa en un bucle con capacidad de predicción. Vamos a argumentar que la consciencia puede emerger a partir de módulos que intentan predecir. Recordemos del capítulo anterior que la inteligencia es principalmente la capacidad de predecir, y suponemos que el robot ya tiene ese nivel de inteligencia. Básicamente, el *yo* es un punto fijo de un conjunto de predictores realimentados sobre sí mismos. Y las sensaciones subjetivas son fallos de codificación. Estas ideas ya las publiqué en (García, 2017 y 2018), pero quisiera aquí mostrarlas más despacio, con más detalles y argumentaciones, cosa que no permiten los artículos.

Para empezar debemos saber que la zona más externa del cerebro humano, llamada neocórtex (de 2 mm de grosor), está conformada por medio millón de columnas corticales de 0,5 mm de diámetro, a su vez organizadas en 6 capas de neuronas. La arquitectura es tan repetitiva que ha llamado la atención de los científicos que trabajan con el cerebro (DeFelipe, 2012). De alguna forma ello está diciendo que el mismo algoritmo que se emplea para percibir sonidos, también se emplea para la visión, para razonamiento espacial, para la lógica y para más cosas. Y ya se sabe cuál es ese algoritmo tan general: cada columna cortical constituye un módulo dedicado a predecir las señales de entrada que le llegan a sus neuronas. Usando una implementación computacional de esta arquitectura, Jeff Hawkins (2004) ha creado un cerebro artificial que está disponible gratuitamente en la web. Y ha cosechado ya algunos éxitos interesantes haciendo, entre otras cosas, que aprenda canciones a base de repetírselas, como nos ocurre a los humanos. En EVALAB hemos usado ese cerebro artificial para que aprenda a cumplir

unas órdenes básicas en un entorno espacial (Narváez, 2016; Narváez, García y Gutiérrez, 2017).

Ahora debemos recordar qué es un punto fijo en matemáticas: si tenemos una función cuya salida se realimenta sobre la entrada, es decir:

$$\begin{aligned} y &= f(x) \\ x &= y \end{aligned} \qquad \text{Ec. 36}$$

Se dice que x_F es un punto fijo de esa función si

$$x_F = f(x_F) \qquad \text{Ec. 37}$$

Para garantizar que exista un punto fijo, la función debe ser continua y contractiva (Wikipedia, 2016b). Y el conjunto de entrada y de salida debe ser el mismo y estar acotado.

Los puntos fijos (Wikipedia, 2016a) son invariantes en un sistema y para entenderlos bien propongo que resolvamos primero el problema 21 de ingenio.

Problema 21: Punto fijo de Palmira a Tuluá

Luisa sale de Palmira a las 8:00 del lunes en su automóvil, y llega a Tuluá a las 13:00. El martes hace el recorrido inverso por la misma carretera, saliendo de Tuluá a las 8:00 y llegando a Palmira a las 13:00. La pregunta es: ¿existe algún punto de la carretera por el que Luisa pasó a la misma hora el lunes y el martes?

Ten en cuenta que no sabemos a qué velocidad circulaba Luisa y que ni siquiera era constante. Seguramente había semáforos, a veces en rojo, a veces en verde. Es posible incluso que en algún momento llevase una velocidad negativa (por ejemplo, si tuvo que dar reversa para recoger algo que olvidó). Lo único que sabemos es que a la ida y al regreso fue por la misma carretera y coincidieron las horas de salida (8:00) y las de llegada al destino (13:00).

La mayoría de la gente piensa que es muy improbable que exista ese punto, pero...

Causa bastante sorpresa este tipo de problemas la primera vez que te enfrentas a ellos. ¿Cómo es posible demostrar que exista algo si no se sabe dónde está? Incluso cuando ya conoces la solución y la has entendido bien, el siguiente problema que te pongan te puede volver a desconcertar (por favor, intenta resolver el problema 22).

Problema 22: Arrugando papeles

Supón que tienes una caja con papel A4 abierta por arriba. Coges el primer papel de todos, lo arrugas de cualquier manera (sin romperlo) y lanzas la pelota de nuevo sobre la caja. Y aquí la pregunta es: ¿existe algún punto del papel arrugado que esté en la misma vertical que estaba antes de arrugarlo?

De nuevo, dado que se puede arrugar de miles de maneras, y se puede lanzar a cualquier sitio encima de la caja, la respuesta es que parece muy improbable que algún punto vaya a permanecer en la misma vertical, pero...

¡Otra vez! ¿Cómo es posible? Parece magia que se pueda llegar a demostrar algo cuando se tiene tan poca información de las condiciones, en este caso, de cómo se arrugó el papel.

Hay una multitud de problemas similares y todos nos causan estupor y asombro. Cada uno plantea algo raro, único y aparentemente desconectado de los demás (Problema 23). Otro interesante y más complicado: busca por Internet el “teorema del no pelo” que demuestra que es imposible peinar por completo una pelota de tenis que tuviera pelo por todos los lados, y que se aplica no a la peluquería sino a algo tan distinto como los agujeros negros en la relatividad general.

Problema 23: La vela

El siguiente problema es para mostrar que los puntos fijos tienen aplicaciones prácticas. Tienes una vela que tarda en arder exactamente una hora. Pero la velocidad con que lo hace no es uniforme pues hay zonas más gruesas que otras. ¿Cómo puedes hacer para cronometrar exactamente media hora?

Volviendo al tema principal, como bien dicen Daniel Dennett y Jeff Hawkins, si tuviéramos un robot con un cerebro sin consciencia pero con percepción sensorial del mundo, las manos, los pies y todos los miembros y sensaciones de su cuerpo formarían parte de su entorno externo. El robot intentaría predecir todo lo que ve y siente, pues si lo logra anticipar podrá salvarse de peligros (hemos supuesto que el robot es el resultado de algún proceso evolutivo, por lo que tendrá cableado en *hardware* todo tipo de algoritmos de supervivencia). También hemos supuesto que el robot ya tiene capacidad para segmentar lo que percibe clasificándolo como objetos de distintos tipos (algo que hacen ya los algoritmos de Google que identifican gatos en videos).

Entonces, el robot se moverá por su mundo, percibiendo objetos y tratando de predecir con mayor o menor acierto lo que van a hacer, cuando de repente ¡algo extraordinario ocurre!: hay un objeto que siempre puede predecir bien al 100%. Si piensa que va a levantar la mano, efectivamente levanta la mano. Si piensa que va a dar un salto, o a hacer un gesto con la cara, efectivamente ello ocurre. Intenta desafiarlo tratando de hacer predicciones sobre movimientos extraños y repentinos⁸⁷ pero siempre acierta. Acaba de descubrir el yo, en su nivel más básico.

⁸⁷ Lo cual seguro que ya habrás experimentado cuando eras un adolescente y estabas delante de un espejo.

Recuadro 8: El "yo" de la consciencia

El yo es un punto fijo que aparece cuando un conjunto de funciones de predicción actúan sobre sí mismas. Y, como tal abstracción matemática, no tiene una existencia material. Es solo una ilusión.

Y causa asombro, claro, como todos los puntos fijos. Y es difícil de entender, y es misterioso... Pero, en fin, es solo eso: un punto fijo.

También es importante entender que mis acciones las decide todo el sistema nervioso incluyendo mi cerebro, mi cerebelo, mi tallo encefálico y otros centros nerviosos. Habitualmente achacamos al *yo* la capacidad de decidir y realizar acciones, ser el origen de la voluntad, ser el conductor que guía al resto del cuerpo, pero ello es incorrecto. Si nos confundimos con ello posiblemente sea debido a que hay una presión, en buena medida evolutiva, por encontrar relaciones de causa-efecto ya que eso facilita la capacidad de predecir el futuro. Y si el cuerpo se mueve mientras ese punto fijo que es el *yo* refleja esos movimientos con total exactitud, es razonable cometer el error de pensar en el *yo* como la causa. Pero los experimentos que menciona Dennett, que ya hemos explicado al principio del capítulo, demuestran que eso no es así. Las decisiones se toman de forma inconsciente por todo el cerebro y el resto del sistema nervioso. La consciencia las registra después de que se han tomado. Parece ser que la consciencia puede tener también una cierta capacidad de veto, es decir, de abortar decisiones que están comenzando a ejecutarse. Pero no decide las acciones.

El hecho de que estos módulos puedan predecir al 100% lo que va a hacer el resto del cuerpo ocurre porque el cerebro maneja información privilegiada: en algún sitio del cerebro y el cerebelo está la información motora, que va a enviarse hacia los músculos. Y conjeturo que los módulos de predicción pueden "hacer trampa" (línea roja de la figura 92), obteniendo de allí la información.

También quiero advertir que, a pesar de ello, no es posible hacer una demostración de la existencia de este punto fijo, pues la función de predicción no es continua (el espacio de objetos sobre la que opera es discreto y muy fraccionado). Suena razonable que un cerebro de robot cuya misión es predecir todo, tarde o temprano se encuentre a sí mismo dentro de ese todo. Pero no hay garantías, pues el cerebro podría tener defectos o peculiaridades que lo impidan. Es decir, pueden existir cerebros sin punto fijo y también cerebros con más de un punto fijo. Ello seguramente está siendo identificado ya como algún tipo de enfermedad, incapacidad o peculiaridad mental.

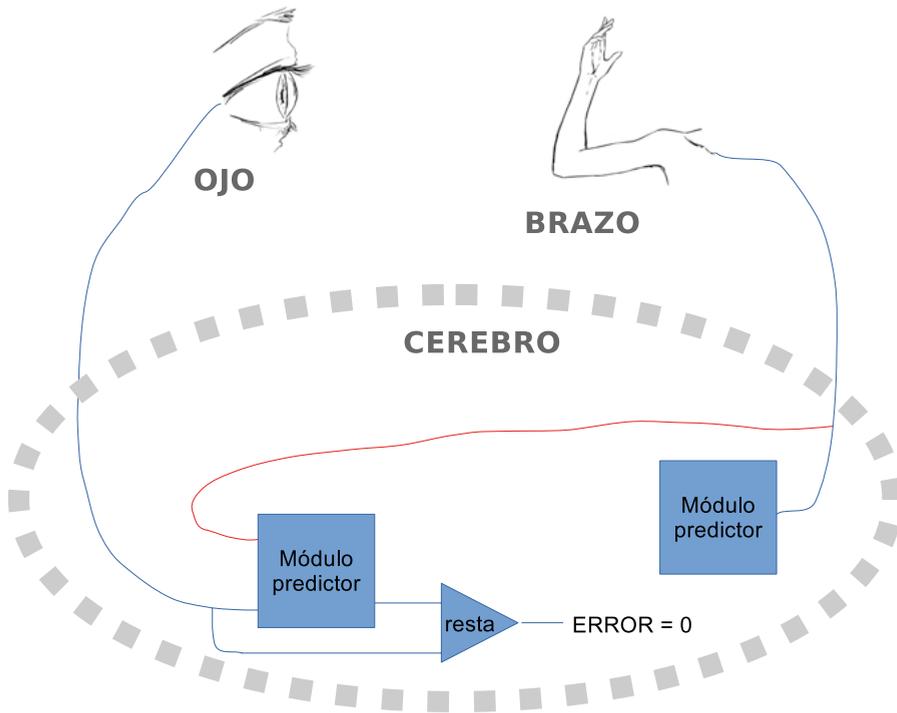


Figura 92. Bosquejo del bucle de predicción con error 0

Es razonable esperar que la presión evolutiva empuje a tener un único *yo* y hay varias razones para ello. Primero, este modelo de mí mismo produce un *yo*, que no solo es el centro de una experiencia narrativa, como dicen Dennett y Bruce Hood (2012), sino que también es a quien debo de salvar en caso de peligro. Ello tiene también obvias ventajas evolutivas. Y segundo, si gracias a la consciencia tengo un modelo de mí mismo (modelo que usa mi mente *popperiana*, como explicábamos en el capítulo anterior), entonces ese modelo también puedo aplicarlo a los otros individuos de mi misma especie, con lo cual puedo predecirlos para saber en quién confiar, con quién puedo cooperar y de quién debo cuidarme. Esta es la hipótesis del psicólogo Nicholas Humphrey (citado por Lewin, 1992, p. 158) acerca de la ventaja de ser consciente en un entorno social. Es lo que se llama tener una teoría de la mente de los otros.

Como novedad quisiera extender la hipótesis de Humphrey en el sentido de que hay una causalidad circular (un bucle de realimentación positiva): al jugar juegos sociales intento predecir a los otros y entonces emerge un objeto especial con el que siempre acierto, el *yo*. Con ello también emerge la autoconsciencia.

Hay un divertido experimento ideado en 1998 por Matthew Botvinick y Jonathan Cohen, de la Universidad Carnegie Mellon y replicado muchas veces (Bernardo, 2013) que corrobora esta propuesta. En este experimento se le pide a una persona que extienda los dos brazos encima de la mesa. El brazo izquierdo queda a la vista mientras que hay una pequeña pantalla que impide a la persona ver su brazo derecho. Al otro lado de la pantalla, y a la vista de la persona, colocan un brazo de plástico. Es obvio que es de plástico y la persona lo sabe. El experimentador se sienta enfrente de la persona, al otro lado de la mesa. Con ayuda de un pincel acaricia la mano de plástico a la vez que, al otro lado de la pantalla y sin que lo vea la persona, con otro pincel acaricia la mano derecha real. La persona siente el cosquilleo en su mano derecha a la vez que ve la mano derecha de plástico ser acariciada con el pincel y de allí llega a algo asombroso: identifica la mano de plástico como si fuera parte del *yo*.

Hay que recalcar esto: no se trata de que la persona sienta un fenómeno curioso de coincidencia sensorial. Es que, al ver la mano de plástico siendo acariciada puede predecir exactamente que va a sentir un cosquilleo; y al sentir el cosquilleo puede predecir con total exactitud que la mano de plástico está siendo acariciada. Recordemos que según esta propuesta, todo lo predecible con 100% de exactitud pasa a denominarse el *yo*, y efectivamente eso es lo que ocurre: la mano de plástico pasa a ser parte del *yo*.

El experimento termina de una manera divertida: el experimentador saca repentinamente un martillo y golpea la mano de plástico. A lo que la persona reacciona visiblemente con susto, gritando o intentando retirar la mano de allí (obviamente solo logra retirar la mano real). Ello corrobora de manera muy efectiva que la persona realmente ha identificado la mano de plástico como una parte de su *yo*, que quiere defender de la agresión.

Hay más experimentos donde al sujeto se le induce a creer que un objeto externo forma parte del *yo* como, por ejemplo, usar un lapicero para tocar objetos de distintas texturas, con los ojos cerrados. Por un lado, es fácil adivinar de qué objetos se trata, a pesar de no estar tocándolos directamente. Y, por otro lado, nos identificamos con la punta del lapicero: nosotros estamos allí, en la punta del lapicero. Lo mismo ocurre cuando nos volvemos expertos en el manejo de una herramienta, y la más cotidiana es el automóvil. Podemos sentir el tipo de terreno que tocan las ruedas (un bache, una piedra...) como si estuviéramos palpándolo con las manos. Podemos también sentir la extensión del automóvil, sabiendo con exactitud si cabe o no cabe entre dos obstáculos. En resumen, si podemos predecir al 100% cómo se va a comportar el automóvil, entonces lo incorporamos al *yo*.

La única crítica seria que he recibido hasta ahora es que los puntos fijos son atractores (o repulsores) de modo que el *yo* debería quedarse atrapado observando al *yo* para siempre. Es muy fácil resolver esta objeción cuando uno se da cuenta que el automóvil del problema 21 tampoco se queda atrapado para siempre en el punto fijo. Hay que distinguir entre el concepto matemático de punto fijo, que es el que estamos tratando aquí, y el concepto en el ámbito de los sistemas dinámicos que, efectivamente, genera atractores, tal y como vimos en el capítulo “Caos” del libro anterior.

Sin embargo, también hay que considerar el cerebro como un sistema dinámico, y sabiendo que un objetivo importante de cualquier proceso de predicción es disminuir su error, entonces el *yo* sí podría convertirse en un atractor en el que, una vez que piensas, no sales jamás.

Esto se puede rebatir. Por un lado, el cerebro no es solo el *yo*. Dado que tiene muchos módulos que trabajan en paralelo, puedo realizar varias tareas a la vez, como conducir un automóvil y escuchar música. Pero definitivamente la consciencia se dirige solo a una cosa a la vez. Hay un circuito de atención (que se puede modelar como en la figura 93), que se centra en uno de los muchos estímulos o pensamientos que lleguen a ese nivel. Y debe haber una funcionalidad que decide qué atender para la que ya hay algunos modelos como el de van den Boogaard y Turpijn (2017). Seguro que tienen prioridad las situaciones de supervivencia. Y si no hay nada de esta índole que atender se podrá conmutar de forma más o menos aleatoria entre todas las posibilidades. En algún momento caeremos en los módulos cuyo error de predicción es cero, pero el conmutador aleatorio nos sacará de allí antes o después. En principio eso no debería generar ningún bucle infinito, salvo que uno se aisle sensorialmente y se concentre mucho en pensar sobre el *yo* o la consciencia. O en alguna situación donde pensar en el *yo* sea vital para algo nuestro (como, por ejemplo, hacer experimentos sobre la percepción del *yo*). Ese tipo de actividades no son aconsejables pues pueden dañarnos la mente, dejándola atrapada, efectivamente, en un bucle de realimentación positiva potencialmente destructivo, en el que cuanto más pensamos en el *yo*, más queremos pensar en el *yo*.

De hecho, los experimentos de privación sensorial (donde se deja flotando al sujeto en un baño a la misma temperatura corporal, sin ningún ruido, luz o estímulos externos), si se prolongan demasiado tiempo pueden producir esos daños en el cerebro, porque el circuito de atención no tiene en qué otra cosa sintonizarse sino en uno mismo. Podemos ver un experimento interesante en Vsauce (2017), que es un excelente canal de divulgación científica.

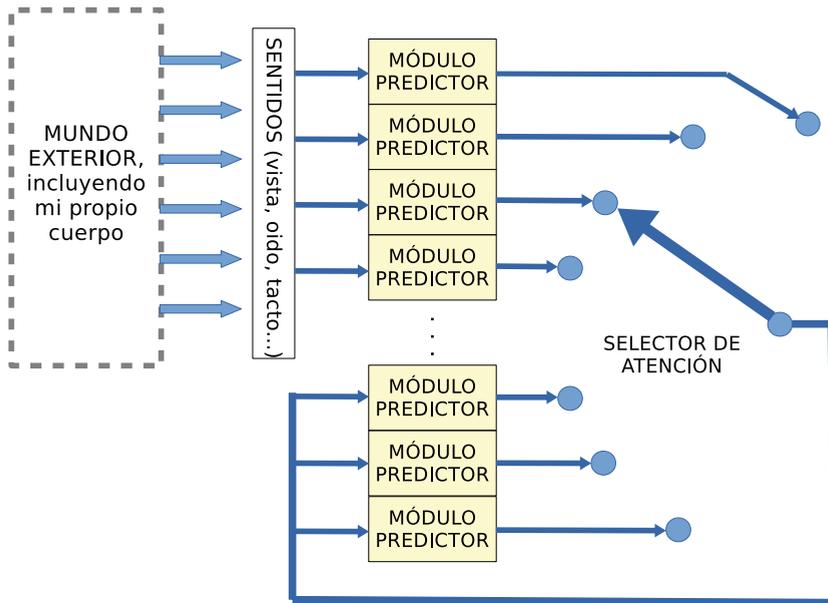


Figura 93. El selector de atención de la consciencia

La idea de que el circuito de atención puede saltar a otras cosas distintas al punto fijo, o sea, que el punto fijo puede desconectarse, se puede intuir a través de una analogía con el proceso de arrugar un papel, que veíamos en el problema 22. Como permitimos que el papel arrugado caiga en cualquier sitio, y hacemos lanzamientos cada cierto tiempo, la mayoría de las veces caerá fuera de la bandeja, y entonces no habrá ningún punto fijo. O sea, la mayoría de las veces la atención del cerebro estará fijada en cosas externas. Pero de vez en cuando el papel arrugado caerá sobre la bandeja y, allí, inevitablemente, aparecerá un punto fijo. Es decir, de vez en cuando pensamos en nosotros mismos, y descubrimos nuestro punto fijo, el *yo*.

En la figura 94 hay un posible diagrama interno de un módulo predictor. La idea de este módulo es que a partir de una secuencia de datos de entrada y de salida se predice lo que va a llegar antes de que llegue. Entonces, si se retrasa la salida, debería coincidir con la entrada. Y si no coincide, un circuito optimizador debe corregir la predicción. Recordemos que en el cerebro tenemos muchas columnas corticales que actúan como módulos de predicción. Y que los *autoencoders* que se usan en *Deep Learning* hacen prácticamente lo mismo: predecir sus propias entradas. De modo que hay bastante coherencia en esta teoría.

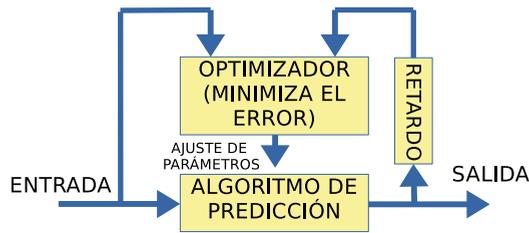


Figura 94. Módulo predictor

El *yo* es un invariante por el que tiene que pasar cualquier sistema inteligente que vaya a ser consciente. Para entender bien este concepto recomiendo volver a recordar el problema de ingenio de salvar el foso del capítulo anterior, junto con el punto por el que hay que pasar. El *yo* es un punto fijo también por esta razón.

Quiero también recalcar qué tipo de argumentación no es esta. No estoy diciendo que como los puntos fijos causan perplejidad, y como la consciencia también causa perplejidad, entonces la consciencia es un punto fijo. Lo que estoy diciendo es que hay un bucle cerrado de predicción que genera un punto fijo y que es sensato identificar ese punto fijo con el *yo*. Y que la noción del *yo* produce perplejidad como la producen todos los puntos fijos.

Evolutivamente hablando, si hay un sistema predictor (Figura 95), por mutaciones se puede cerrar el bucle, con lo cual emerge el concepto del *yo* (Figura 96).



Figura 95. El cerebro, prediciendo lo que percibe y tomando acciones

La consciencia no termina aquí. Más bien, comienza aquí, con este bucle. Desde un punto de vista ingenieril, es este bucle el que crea la ilusión del *yo*, que se convierte en el sujeto de la narrativa de las sensaciones que recibo como entradas o como estados internos, y el aparente sujeto causal de las acciones que realizo. En ese primer nivel “predigo lo que voy a hacer y hago lo que estoy prediciendo”. Pero no puede olvidarse que el cerebro humano es muy complejo, de modo que es de esperar que haya muchos otros bucles actuando en varios niveles de descripción, en paralelo, y con unos sirviendo como metáfora, entradas o regulaciones de otros. Por ejemplo, al involucrar los sentidos “veo que mis manos se mueven como ordené, a la vez que hago mover las manos como lo predije”. Todos esos bucles hacen referencia implícita al *yo*. Crean el *yo*. Y ello construye nuestra rica experiencia consciente.

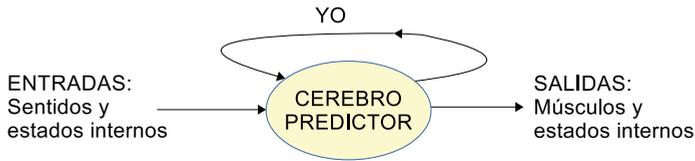


Figura 96. En el cerebro se forma un nuevo bucle

Además, si recordamos que las mentes *popperianas* contienen un modelo del mundo implementado como un simulador, desde esta perspectiva también emerge un punto fijo: dentro del mundo hay un objeto que simula el mundo. Se trata de otra función contractiva que se mapea sobre sí misma, que contendrá el mismo punto fijo: el yo.

Como mencionábamos en los experimentos al inicio de este capítulo, la consciencia es muchas veces un observador, y no quien dirige la acción. En ellos se ha detectado un retraso de tiempo positivo entre la acción y la consciencia de la acción. En este modelo que propongo, ello surge naturalmente. El *yo* es solo un punto fijo, sin existencia como objeto. Es la correspondencia entre lo que predigo que voy a hacer y lo que hace ese cuerpo, que resulta que es el mío. Los módulos de predicción vaticinan que va a ocurrir una acción en mi cuerpo porque las salidas neuronales para desencadenar esa acción ya han ocurrido. Por cuestiones de causalidad, esta predicción sucede unas décimas de segundo después de que la acción ya se haya decidido a nivel inconsciente. A su vez, como somos incapaces de detectar por nosotros mismos esa diferencia de tiempos tan pequeña, reinterpretamos la situación como que el *yo* ha decidido realizar una acción.

Desde mi punto de vista, eso no significa que sea imposible tomar decisiones conscientes. Lo que significa es que es imposible hacerlo en el mismo momento en que se debe reaccionar rápidamente. Pero cuando hay suficiente tiempo disponible, la consciencia sí puede tomar decisiones, básicamente creando situaciones reales o, la mayoría de las veces imaginarias (usando nuestro simulador *popperiano*), donde entrenar al inconsciente para que en un futuro enfrente esas situaciones de cierta manera.

De la misma forma, un guardameta de fútbol se entrena *offline*, que es cuando tiene suficiente tiempo, para crear situaciones que probablemente se van a dar durante un partido real, adiestrando así sus músculos, sus sensores y su sistema de decisión inconsciente. Ya que cuando tenga que detener un *penalty* real, la consciencia no va a tener tiempo para tomar decisiones elaboradas. Un ejemplo similar es cuando una joven decide entrar a la universidad. En un futuro, durante su vida profesional no va a tener tiempo de ponerse a estudiar usando libros para resolver cada problema que se le presente. Pero el entrenamiento previo ya lo hizo en las aulas, recabando infor-

mación y adiestrando habilidades. Es allí donde interviene la consciencia, en modo *offline* entrenando el inconsciente para situaciones futuras. Y ello tiene una consecuencia filosófica y jurídica: definitivamente sí somos responsables de nuestras decisiones, así sea porque no nos entrenamos para afrontar los problemas (estudiando o practicando), cuando había tiempo para ello.

El yo es solo un aspecto de la consciencia, pero está en el núcleo de todo lo demás. Con esta propuesta se abre un camino para que la inteligencia artificial tenga un yo, con todas las consecuencias positivas y negativas que pueda haber. Sin embargo, queda un obstáculo bastante filosófico y es que actualmente no disponemos de *tests* de consciencia. Hay muchos *tests* de inteligencia como IQ-test.cc (2017), 123test (2017), IQ_Research (2017) o Mensa (2017) que miden el IQ⁸⁸ de una u otra manera. Son bastante discutibles porque contienen sesgos culturales, idiomáticos y otros, pero algo indican. Sin embargo, no hay nada parecido para medir la consciencia. Y es importante tener *tests* de consciencia, para poder constatar los progresos que haga la inteligencia artificial en este sentido.

No obstante, hay un indicador de consciencia, quizás el único hasta ahora. Se trata de poner al sujeto bajo *test* delante de un espejo y mirar si se reconoce a sí mismo, o dicho en mis términos, si puede predecir lo que hace la imagen en el espejo con error igual a cero. Esta prueba la han pasado con éxito las siguientes especies: los grandes simios (humanos, chimpancés, gorilas, bonobos y orangutanes), los delfines, las orcas, los elefantes asiáticos, algunas hormigas y un pájaro llamado *magpie* según recogen Pachniewska (2017) y Think_Elephants_International (2015). No tengo constancia de que se haya hecho la prueba con pulpos ni con cuervos, que también tienen mucha inteligencia. Pensemos que aunque una especie haya pasado el *test* no significa que todos sus miembros lo hagan. Hay algunos individuos, típicamente los más fuertes, que simplemente muestran comportamientos agresivos frente al espejo.

La primera fase de un animal al verse ante un espejo es de comportamiento social. Creen estar ante otro individuo y reaccionan con amistad, agresividad o indiferencia. La mayoría de los animales como perros, gatos y peces se comportan agresivamente y nunca salen de esa fase. La segunda fase es verificar el comportamiento. Intentan distintas posturas y movimientos para comprobar si la imagen los imita. La tercera fase es de comprensión de que esa imagen es de uno mismo. Allí prueban a mirarse partes de su cuerpo que nunca antes habían podido ver, e incluso a coquetear delante del espejo. La cuarta fase la imponen los investigadores humanos para verificar

88 *Intelligence quotient.*

todo lo anterior. Para ello colocan una marca claramente visible en el animal. Pero el animal no la puede ver, excepto por medio del espejo. Cuando el animal la descubre, si se interesa por la marca en la imagen, ello indica que no ha comprendido nada. Solo si se interesa por la marca en el propio cuerpo es que ha pasado esta prueba de consciencia.

EL “HARD PROBLEM”

Con ello queda explicado el *yo* pero todavía hay un problema aparentemente más misterioso: la percepción subjetiva de la realidad, el flujo continuo de sensaciones que experimentamos, los *qualia* que memorizamos y recreamos en nuestro interior, en fin, lo que Chalmers llama el problema difícil⁸⁹ de la consciencia. Solo que él lo resuelve de una forma un tanto esotérica, suponiendo que todos los objetos del universo tienen consciencia y que ella es, por así decir, la quinta fuerza fundamental, en lo que se ha dado en llamar *pansiquismo*, también defendido por el filósofo Christof Koch (2017), que cree que la inteligencia y la consciencia son cosas completamente distintas. Muy al contrario, en este libro abogamos que la segunda es una consecuencia de la primera y que ambas son fenómenos computacionales. Chalmers también opina que es imposible que los computadores logren consciencia, porque sigue pensando que no es lo mismo una simulación de una tormenta, que una tormenta real. La última moja. Pero eso es absurdo por dos razones. Primera: para los seres que vivan en el mundo de la tormenta simulada, la lluvia simulada también moja. Segunda: si consideras que nuestro mundo es solo información, que se procesa de múltiples formas, y sabes que cualquier máquina universal de Turing puede simular a cualquier otra, entonces, no hay diferencia entre vida, inteligencia y consciencia reales versus simuladas.

Al contrario que ellos, Dennett niega que existan esos *qualia* y la experiencia subjetiva que conllevan. Aquí voy a proponer una solución a este dilema, muy cercana a la de Dennett, pero partiendo de bases meramente computacionales (García, 2018).

Para empezar debo decir que estoy de acuerdo en que la experiencia interna del color rojo o de sentir mucha hambre, no es lo mismo que si un robot en su *software* tiene un par de variables como:

color= RGB (255,0,0)

hambre= 95%

Ec. 38

No es lo mismo, pero... sí. En el fondo es lo mismo. Me explicaré.

89 *Hard problem.*

Sabemos que no hay un homúnculo⁹⁰ dentro del cerebro, o sea, no hay ningún control central. Por el contrario, sabemos que la información se encuentra allí distribuida en grandes redes neuronales. De modo que definitivamente no hay variables con valores asignados como en la ecuación 38. Pero nada impide que haya algo similar distribuido, usando algo así como memoria asociativa (parecida a la que se emplea en las *cachés* de los computadores). El problema no es la implementación, que puede ser cualquiera. Sea como sea, la información de la ecuación 38 está allí y los neurocientíficos no tardarán mucho en explicarnos cómo se almacena en el cerebro.

Pensemos primero si los robots van a tener *qualia*. Un robot como los que fabricamos actualmente va a tener variables como las de la ecuación 38 porque emplearemos algún lenguaje de programación estándar, y todos ellos tienen variables, pues es la forma más obvia de representar información. Y si a un robot le preguntamos por su estado interno, por sus *qualia*, responderá con toda precisión “estoy viendo que, al esconderse el sol, se produce un fenómeno de difracción de luz en la atmósfera cuyo color predominante se representa con el código RGB 255,0,0 que cae en la franja de lo que vosotros humanos llamáis rojo; y mis baterías se han agotado en un 95% por lo que necesito recargarlas” al más puro estilo de Mr. Spock de *Viaje a las estrellas* (Roddenberry, 1966) o de Sheldon Cooper en *The Big Bang Theory* (Lorre y Prady, 2007). Desde luego que no está sintiendo hambre y no está experimentando el esplendoroso atardecer de la figura 97, sino que solo puede dar detalles de sus estados internos de manera fría y objetiva.



Figura 97. Atardecer en Salvador, Bahía

90 Si crees que dentro de tu cerebro hay un pequeño personaje tomando tus decisiones, entonces dentro del cerebro de ese personaje debe haber otro, y luego otro y otro... lo cual no tiene mucho sentido. El cerebro es un sistema distribuido de neuronas sin control central, del mismo modo que veíamos que los hormigueros y bandadas de aves no tienen un jefe que las guíe. En las personas más extrovertidas, una parte del sistema distribuido es externo al individuo, que no puede tomar decisiones o ni siquiera existir con cordura sin contar con la red de apoyo de sus amigos.

Hacer que el robot no solo pueda reportar el valor de descarga de sus baterías, sino que también sienta realmente hambre es bastante sencillo. Teniendo en cuenta, como dijimos, que fue creado por medio de un proceso evolutivo, eso significa que él sobrevivió, mientras que todos aquellos robots que no supieron buscar una fuente de energía cuando sus baterías estaban próximas a descargarse murieron en el proceso y no dejaron descendientes. Por ello, es de esperar que este robot cuando vea una fuente de energía se detenga a observarla, a sopesar cual es el nivel de batería que todavía dispone, a calcular si es suficiente y si puede llegar a otra fuente de energía. Conforme la batería esté baja, el robot comenzará a distraerse de sus tareas habituales, buscando con afán una fuente de energía e incluso corriendo hacia ella cuando su propio nivel esté en un punto crítico, olvidando sus otras obligaciones. Sus ojos se le irán hacia allá y todos sus ademanes delatarían la tensión interna de vida o muerte a la que se encuentra sometido. Cualquier observador diría que está sintiendo hambre, pues muestra todos los síntomas. La evolución hizo el trabajo de cablear en *hardware*, por motivos de supervivencia, unas respuestas que muestran intencionalidad⁹¹, ganas de acercarse al enchufe, vaya, está sintiendo hambre de energía. Si quieres saber cómo la evolución dota de intencionalidad a un robot, puedes ver el video de Karl Sims (1994), especialmente en la sección donde el objetivo es seguir un punto rojo. Se trata de un robot que aprende a nadar y a seguir un punto rojo por medio de algoritmos evolutivos, después de muchas generaciones y sin que ningún programador humano haya escrito ni una línea de código para dirigir su comportamiento. Todo fue aprendido en la larga lucha por sobrevivir, pues a los robots que lograban acercarse más rápidamente al punto rojo se les permitía tener más hijos. Al ver este video, muchos de mis estudiantes dicen que al robot *le da rabia* cuando el punto rojo se aleja, y vuelve a nadar vigorosamente hacia él. Dicen que el robot tiene un propósito claro, que tiene intenciones de seguir al punto rojo a donde quiera que vaya. Podemos decir que su estado interno, la percepción de que el punto rojo se aleja de él, le suscita emociones imposibles de controlar, que le empujan a seguirlo. La evolución cablea⁹² esas emociones.

Volviendo al ejemplo del robot hambriento (basta con interpretar el punto rojo como una fuente de energía) es importante entender esto en toda su profundidad: si le preguntamos al robot por qué se afana en aproximarse a la fuente de energía en vez de continuar el trabajo que estaba realizando nos

91 Intencionalidad en el sentido de mostrar intenciones de hacer algo. La definición filosófica de intencionalidad abarca este concepto pero es bastante más compleja.

92 Cablear es un concepto electrónico que significa que no es programable, que nacimos así con ello.

dirá que no lo sabe, que es algo más fuerte que él, que no lo puede controlar. Y ello es debido a que los comportamientos encaminados a lograr la supervivencia tienen prioridad sobre cualquier otro comportamiento. Los robots donde esto no es así hace tiempo dejaron de existir, por razones evolutivas.

El problema grave está con los *qualia* que no producen efectos observables, como regocijarse ante una puesta de sol. La evolución aquí no presiona en ningún sentido pues no se trata de una cuestión de supervivencia. Quiero matizar esto. Desde luego que poder diferenciar colores sí es importante para sobrevivir porque permite distinguir mejor a animales peligrosos, pero que esa combinación de colores me produzca más emoción estética que aquella no sirve para nada práctico. No tiene ningún valor evolutivo orientado a la supervivencia o reproducción.

Entonces aunque la representación de los colores en nuestra mente si puede ser el resultado de un proceso evolutivo, el placer estético que sentimos al verlos o recuperarlos de nuestras memorias no equivale a un frío código 255,0,0. La sensación subjetiva del color y, en general, de todos los *qualia* que no producen efectos observables, no tiene una explicación evolutiva.

La explicación de esos *qualia* viene de otro lado: se trata únicamente de un problema de codificación. Los *qualia* contienen información sensorial del mundo y de nuestros estados internos, pero esa información está encriptada en cierto modo. Podemos recuperarla, podemos trabajar con ella, pero no podemos entenderla en su detalle.

Ello se debe a que el *hardware* de un robot es muy distinto al *wetware* de los cerebros biológicos. En los robots la información está codificada en binario y desde allí se puede traducir biunívocamente a cualquier otro código numérico que represente números en decimal, letras, colores y sonidos. Pero la representación básica es binario. Mientras que en los cerebros biológicos la información básica es difusa y está representada por disparos eléctricos de neuronas. Y cualquier otra información (sonidos, letras, colores...) se soporta directamente en ese código, como conjuntos y secuencias de disparos eléctricos, a lo que se le suele llamar patrones de actividad neuronal. Además, las representaciones no son únicas y esos patrones se intersecan unos con otros a la vez que están distribuidos por amplios conjuntos de neuronas. Es decir, una neurona que dispara de cierta forma cuando la consciencia experimenta el color rojo, también puede estar disparándose de forma similar cuando la consciencia detecta un sonido particular. Los patrones son distintos y las experiencias que representan también, porque el conjunto de neuronas involucrado es distinto, pero puede haber algunas intersecciones.

Por ello no es de extrañar que ciertos colores o ciertos acordes menores nos pongan tristes: hay una intersección en los patrones de tristeza y de

música en acorde menor. Algo similar pasa con las representaciones computacionales en matrices dispersas que usa Hawkins en su cerebro artificial NuPic: hay una pequeña probabilidad de colisión entre símbolos distintos. Además, en el córtex una señal eléctrica que represente un sonido y otra que represente una imagen son indistinguibles por sí solas (Hawkins y Blackeslee, 2004). Son patrones que no están etiquetados. La diferencia está en el contexto, en las relaciones con otros patrones, de la misma manera que en un grafo dos nodos pueden parecer muy similares en cuanto a sus patrones de conexiones, pero los nodos vecinos son distintos y uno se refiere al sonido mientras que el otro lo hace a la visión.

En resumen, el *wetware* neuronal no es tan preciso como el *hardware* digital⁹³. ¿Por qué se dan esas interferencias y no otras? ¿Por qué al oír un acorde menor me pongo triste en vez de sentir hambre? Es algo accidental, aleatorio, que está embebido en nuestra arquitectura cerebral y que se irá descubriendo en detalle conforme se obtenga el mapa completo del cerebro. Pero es absolutamente accidental pues no ha habido allí presiones evolutivas para que ocurra de esa manera. En realidad sí las ha habido en el sentido de que las interferencias nunca son buenas. Pueden ser neutras (las que ocurren entre estados totalmente internos del cerebro, como música, color y tristeza) y pueden ser perjudiciales (las que producen comportamientos externos observables, como sería ponerse a correr al ver el color azul y paralizarse al ver el color amarillo). Las perjudiciales han tenido que sufrir una selección adversa y, por ello, no es de esperar que queden muchas ni en mucha intensidad, pues quien las porte sería presa fácil de manipulaciones que probablemente acaben con su vida, con lo cual no dejará muchos hijos que sigan portando tales anomalías.

Retomando el hilo, la codificación de los *qualia* es no numérica, distribuida, difusa y con una cierta ambigüedad. Está encriptada. Por “encriptada” no me refiero a que use una codificación fuerte tipo RSA de 1024 bits, sino simplemente que no hay una relación obvia, lineal y directa entre la entrada y la salida. Y tampoco hay una función inversa que me recupere una entrada a partir de una salida.

Analicemos un poco más los *qualia* del color. Podemos ignorar la intensidad de la luz, pues la verdad es que está codificada de una manera bastante lineal en nuestros *qualia*: más luz externa produce más sensación interna. Centrémonos únicamente en el color, que debería ser una única dimensión: la longitud de onda. Sin embargo, debido a que la evolución biológica es

93 Y, precisamente por ello, es más robusto el cerebro. Si unas neuronas se mueren no pasa nada, pero si un cable de computador se rompe, todo el computador deja de funcionar.

bastante chapucera⁹⁴ y optimiza solo en el corto plazo, los ojos humanos terminamos con tres sensores, cada uno de ellos centrado en una longitud de onda distinta, aproximadamente en el rojo, verde y azul. Los ojos entregan información de la diferencia de la longitud de onda de la luz incidente respecto al centro de la banda de cada sensor, y lo que debería ser una única dimensión ahora son tres. Ese mundo de colores en el que estamos sumergidos es solo una ilusión. Son *qualia* que codifican de una forma retorcida una información externa lineal.

Hemos diseñado las cámaras que usan nuestros robots repitiendo ese mismo esquema, y lo llamamos codificación RGB, que usa tres números. Pero nada impediría, en principio, hacer una cámara que entregue por cada pixel un único número: la longitud de onda incidente, en nanómetros. Al respecto, podemos ver una tabla de traducción con algunos códigos en la figura 98.

Longitud de onda (nm)	Código RGB	Nombre
380	97,0,97	Morado
440	0,0,255	Azul
486	0,239,255	Azul aguamarina
558	189,255,0	Verde limón
580	255,255,0	Amarillo
614	255,141,0	Anaranjado
699	255,0,0	Rojo
764	132,0,0	Rojo granate

Figura 98. Longitud de onda de la luz, código RGB y nombre convenido.
Adaptado de García (2018)

Si construimos un robot con una cámara así, y le pido que mire una manzana cuyo verde corresponde a 558 nm, y mañana le pregunto de qué color era, irá a su memoria, recuperará el color almacenado y me dirá que es 558 nm. Almacenó el número 558 y recuperó el número 558 (en la figura 99 podemos ver el correspondiente diagrama de bloques supersimplificado). Le pregunto después si su experiencia con los colores es tan maravillosa como la mía. El robot no lo puede saber, porque mis *qualia* de color permanecen

⁹⁴ El *bricoleur* de Jacobs.

confinados en mi mente. Pero me dirá que su experiencia de color no tiene nada de maravillosa sino que se rige por la lógica: almacenó 558 y recuperó 558. El robot no disfruta de *qualia* de color en el mismo sentido que tenemos los humanos.

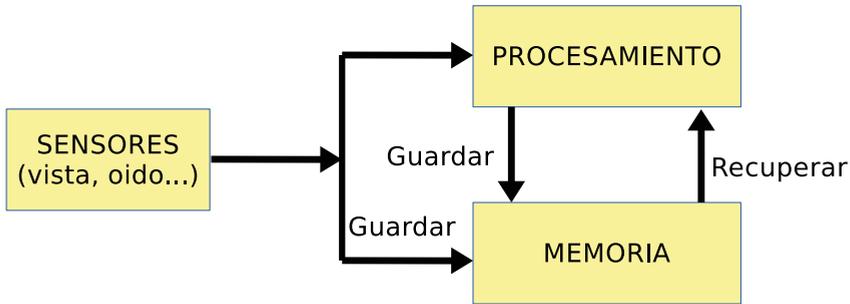


Figura 99. Bloques internos de un robot sin qualia
Adaptado de García (2018).

Es más, al ser tan precisa su representación interna de datos, puede externalizarlos sin que se pierda nada. El robot si tiene función inversa para cada *quale*. Y puede compararlos con los de otros robots, cosa que nosotros los humanos no podemos. Por ejemplo, no podemos llegar a determinar si el rojo que yo vi es más bonito que el que tú viste, y la razón de ello es que esa encriptación de datos nos lo impide.

Ahora hagamos algo maquiavélico con el pobre robot. Cuando esté dormido y sin que se dé cuenta, añadamos un módulo de encriptación de la información, justo antes de que se procese o almacene en memoria, de la forma que se indica en la figura 100.

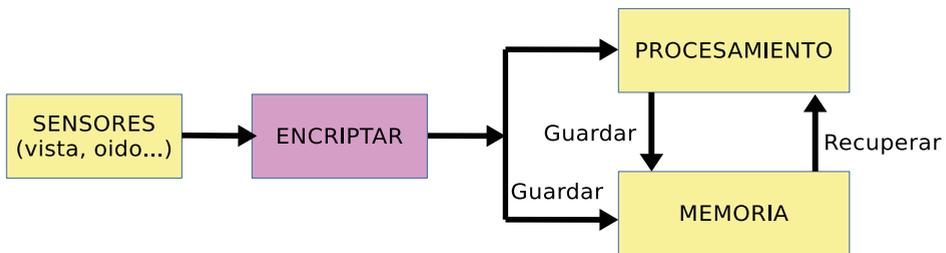


Figura 100. Bloques internos de un robot con qualia
Adaptado de García (2018).

Tomemos también todos los datos previamente almacenados en su memoria (de experiencias que haya tenido en el pasado) y encriptémoslos, para que no pueda notar discrepancias con la forma antigua de codificar,

respecto a esta nueva. Cuando vea una pera verde y la compare con otra pera verde que vio hace unos días y que mantiene en memoria, ambas coincidirán aproximadamente. Pero ahora ya no puede hacer comparaciones con los colores que están viendo otros robots, pues ya no tiene función inversa para traducir un estado interno de su memoria a un valor sensorial externo. Para hacer el símil más exacto, deberíamos encriptar usando un código no numérico, quizás como un conjunto borroso de trazos (Figura 101) o algo similar. Entonces el robot se sentirá aturdido pues ve un color, en su memoria tiene ese color, pero no es capaz de expresar hacia afuera qué es y mucho menos compararlo con el de otros robots. El robot no podrá por menos que maravillarse, y hacer conjeturas sobre esto nuevo que está experimentando. ¡Acabamos de implementar *qualia* humanos en nuestro robot!



Figura 101. Patrones de información distribuidos, no comparables entre sí

Mi propuesta es que los *qualia* en humanos y la experiencia subjetiva que emana de allí son solo eso: datos codificados de una manera no numérica, no lineal y no invertible. Datos encriptados. Y a eso es lo que llamamos experiencia subjetiva, porque no se puede traducir a nada conocido⁹⁵. Y, sabiéndolo así, nos damos cuenta que los *qualia* no tienen mayor trascendencia, ni sirven para nada práctico. Es más, si hubiera suficiente presión selectiva, la evolución se encargaría de eliminarlos.

En el caso de los humanos, la encriptación ocurre directamente en el sensor. Por ejemplo, nuestros ojos nos dan la información en un formato raro de 3 variables (RGB) y además por medio de patrones de pulsos eléctricos.

Llamamos “sensación subjetiva” a aquella que no podemos objetivar, es decir, no podemos sacarla hacia afuera, para que la vean los demás. La “sensación de amarillo” entra encriptada en nuestra mente, donde se representa por un código único que es imposible volver a convertir hacia el exterior. Es decir, no tiene función inversa y por ello los códigos únicos se nos antojan tan misteriosos.

No es solo que la información va perdiendo detalles y se hace más abstracta conforme atraviesa más capas neuronales de nuestro cerebro. Es que

⁹⁵ Esto es a lo que David Chalmers llama el doble aspecto de la información, solo que mi propuesta lo despoja de misterio y lo plantea en términos computacionales.

realmente no hay como saber de dónde vino la sensación inicial. Hemos visto tres árboles, y podemos meditar con el número tres, que es una abstracción, y así pensamos en que es un número impar y primo. Pero más adelante podemos volver a imaginar los tres árboles sin ningún esfuerzo. La información original es recuperable. El problema con los colores es distinto. Hay una pérdida irrecuperable de información que nos impide volver atrás. El mismo color verde de los árboles lo podemos recrear en nuestra imaginación, lo podemos pintar en un cuadro y ajustar la paleta hasta que sea muy parecido al que tenemos en la memoria. Pero no sabemos de dónde viene ese color. No tenemos la noción de longitud de onda en ninguna parte de nuestra imbricada red neuronal. Y tampoco tenemos ninguna otra noción linealmente relacionada con la longitud de onda. La longitud de onda es un escalar mientras que el color es un vector de tres dimensiones dentro de nuestro cerebro, y la relación entre ambos es no lineal e incluso con ambigüedades. Eso es lo que nos impide entender el color.

También podemos plantear un experimento mental inverso al del robot: eliminar la magia de los *qualia* en los humanos. Para ello, con ayuda de equipo externo informamos a una persona de las longitudes de onda que está viendo en cada momento, y le pedimos que las memorice repitiendo el proceso muchas veces. Aunque a bajo nivel (neuronal) no existen los números, a alto nivel (simbólico) si los podemos manipular y memorizar en nuestro cerebro. De modo que el sujeto bajo pruebas, a la vez que memoriza el azul de la forma habitual, también memoriza 440 nm. Después de entrenarlo un rato con muchos colores le podemos mostrar un color concreto y pedirle que nos diga cuál es su longitud de onda. Lo hará sin sentir ningún misterio acerca de *qualia*.

Por último, pensemos en otro tipo de información interna, relacionada con la visión: gracias principalmente a que tenemos dos ojos, podemos triangular y averiguar la distancia aproximada a que se encuentra cada objeto. Es información interna, son también *qualia*, pero nadie se maravilla de ellos como ocurre con el color. Nadie dice “¡Uy!, qué distancia lejana tan hermosa” o “¡qué distancia intermedia tan especial!” mientras que sí lo hacemos con los colores. ¿Por qué? La respuesta es sencilla y viene a corroborar lo dicho hasta ahora: en las distancias sí ha habido presión evolutiva para que la codificación sea lineal, con el objetivo de permitir comparaciones. Es muy importante, por razones de supervivencia, saber si el tigre que se aproxima está muy cerca o muy lejos, o si el árbol al que me voy a subir está más cerca que el tigre. Mientras que no tiene ninguna importancia saber si

las hojas del árbol son más amarillas que la piel del tigre⁹⁶. Como resultado, la codificación de colores está arbitrariamente encriptada mientras que la codificación de distancias no lo está.

El hecho de que los *qualia* mantengan información privada desconectada del mundo real, aunado a que hay interferencia entre *qualia*, todo ello genera confusiones que interpretamos de una forma poética. Tenemos experiencias subjetivas. Pero si somos más rigurosos, lo que esto nos dice es que la evolución no ha trabajado todavía suficiente, seguramente porque no tenía incentivos para hacerlo. Nuestros cerebros son poco evolucionados. Puede haber cerebros mejores.

Un simulador, como el que propone Dennet que tienen las mentes popperianas, es un objeto computacional. De hecho, su formalismo corresponde a un computador. Y la forma más efectiva de controlar un computador es por medio de un lenguaje de programación. De manera que propongo una nueva utilidad de los lenguajes humanos. No solo sirven para comunicarse unas personas con otras, sino también cada persona con su simulador, o dicho más directamente, cada persona consigo misma. Las presiones evolutivas empujaron a crear lenguajes por esta doble ganancia. Y con ello los humanos (y cualquier animal que vaya adquiriendo un lenguaje) vamos abandonando el mundo de las experiencias subjetivas para alcanzar un grado más alto de consciencia. Experimentamos el color amarillo, podemos recuperar una vívida impresión de él si así lo deseamos, pero la mayor parte del tiempo simplemente usamos como sustituto la palabra “amarillo”, porque es más concreta y flexible, se le pueden añadir otras palabras para modificarla y matizarla. El uso del lenguaje permite un control más preciso de nuestro respectivo simulador y también permite comunicarse con otras personas. En este sentido, los *qualia* y la subjetividad son lo contrario de la consciencia y se han interpretado mal hasta ahora.

Estoy recordando un momento muy gracioso de la película *Words and Pictures* (2013) donde están enfrentados un profesor de lenguaje y una profesora de pintura. La profesora de pintura alega que una imagen vale más que mil palabras. El profesor de lenguaje pide a sus estudiantes que expliquen visualmente algunas ideas, tales como “estar hambriento” o “me gustan”, lo cual los estudiantes consiguen a base de gestos y gruñidos. Pero luego les pide que hagan lo mismo con “reúne a la tribu esta tarde que quiero explicarles que ya casi no hay comida y que tendremos que decidir hacia

96 Obviamente, sí tiene importancia el color para el humano, si desea descubrir al tigre camuflado entre hojas amarillas. Pero la presión evolutiva empujaría en este caso a lograr una mejor resolución en el sensor de color, donde se distingan más matices de amarillo. Y no a linealizarlo.

dónde movernos”. Los estudiantes quedan perplejos, pues es imposible visualmente transmitir esa idea. Con ello queda demostrada la superioridad y riqueza del lenguaje frente a una mera impresión visual; de la consciencia soportada por el lenguaje frente a la mera experiencia subjetiva.

Al respecto recordemos que en el capítulo anterior se mencionaban dos métodos de resolver problemas que empleamos los humanos: el intuitivo, que se corresponde con el mundo de las experiencias subjetivas; y el razonamiento estructurado, que es el que necesita un lenguaje para poder ser llevado a cabo.

Para terminar, quiero comentar que un átomo no es consciente en algún mínimo grado, como indican Chalmers (1995) y Tononi (2015). El cerebro es una acumulación organizada de átomos, pero la consciencia no es una suma de pequeñas consciencias atómicas. Analicemos esto con atención. El peso del cerebro es simplemente la suma de los pesos de los átomos, mientras que la consciencia del cerebro emerge debido a la interacción de un enorme número de neuronas. Cuando un átomo recibe las vibraciones mecánicas de una música escrita en acordes menores no es que sienta una tristeza muy, muy, muy pequeña. Es que no siente ninguna. El concepto de tristeza no existe en las partículas elementales. No ha emergido todavía. El fotón no tiene intencionalidad, ni sentiencia, ni posee *qualia*. A lo sumo, quizás tenga algo de libertad, en el sentido explicado en el capítulo del mismo nombre.

LA CONSCIENCIA ES LA BUROCRACIA DEL CEREBRO

No hagas hoy lo que puedas dejar para mañana, pues a lo mejor mañana ya no hay que hacerlo,

CLARA E. GARCÍA BAÑOS

Hay aspectos del *yo* que son bien conocidos y que comparto plenamente. En particular, el *yo* no decide nada. No es quien está a cargo del cerebro y del cuerpo del individuo. No es quien goza de libertad, como discutíamos hace un rato. Y ya hemos visto las interpretaciones computacionales. Pero no quiero terminar aquí, pues hay otra forma de ver la consciencia y que nos puede servir sobre todo para entender por qué surgió. Sabemos que es lenta. Solo observa. Si toma acciones no son inmediatas sino de largo plazo. Se parece mucho a la burocracia de cualquier organización, y vamos a ver sus semejanzas.

Primero comencemos explicando la ley de Parkinson, que es empírica y que fue formulada por un funcionario inglés del mismo nombre en 1955, en India, mientras intentaba comprender por qué en la Oficina de la Colonia

Británica requerían cada vez más empleados. Se enuncia de muchas maneras. Una es así: “cualquier trabajo ocupa todos los recursos disponibles”. Parkinson se dio cuenta que el personal crecía exponencialmente, a una tasa de alrededor de un 5% anual, independientemente de que se logaran mayores o menores producciones de té. Y las razones que esgrimió para ello eran dos, que las voy a formular en términos modernos: la primera es que cada funcionario a lo largo de los años se cansaba de su trabajo y entonces contrataba a otros dos, más jóvenes, para que lo hicieran. De este modo repartía el trabajo entre los dos, pero impidiendo la comunicación directa entre ellos, con el objetivo de mantenerse imprescindible y evitar que lo despidieran. Nadie sabía todo sobre su trabajo, y todas las comunicaciones y decisiones pasaban por él. La segunda razón es que la gerencia quería saber el estado de la organización, pidiendo informes a los empleados que a su vez, para poder escribirlos, pedían informes a otros empleados, generando más y más trabajo.

Con ello aumenta exponencialmente la burocracia: las unidades internas se ponen tareas unas a otras, aun cuando no produzcan nada hacia el exterior. Y contratan más personal para poder hacer esas tareas agobiantes. En este sentido, la emergencia de la consciencia parece inevitable una vez que tienes muchos agentes encargados de realizar tareas concretas, usando los datos generados por otros agentes o el entorno, y ofreciendo los resultados al entorno o a otros agentes.

En el cerebro algo está pasando que se parece bastante, pues aunque constituye el 2% del peso del cuerpo, consume aproximadamente el 20% de sus recursos energéticos. Recordemos además que la consciencia no toma decisiones sino que mira lo que ocurre y lo anota, y discute consigo misma sobre los porqués, y compara con situaciones pasadas y se deprime, y en la confusión decide no actuar cuando podría hacerlo. Como Aureliano Buendía⁹⁷, la consciencia registra y pone etiquetas a todo. Y como la oficina que investigó Parkinson, se inventa tareas que ponerse internamente, unos módulos a otros, sin ejecutar hacia afuera ninguna acción.

Por cierto que la ley de Parkinson tiene corolarios en muchos ámbitos, incluida la computación: “el sistema operativo usará todos los recursos disponibles”; o “los discos duros se llenan por muy grandes que sean”. Pero también tiene una aplicación bastante interesante para administrar el tiempo: “si lo dejas para el último minuto, te llevará solo un minuto hacerlo” y eso muestra que la procrastinación no es tan mala como todo el mundo cree,

97 Ver cita del principio del capítulo.

en consonancia con la cita del principio de este apartado. La procrastinación puede ser pereza, pero también puede ser una estrategia de optimización.

Hay otra conclusión similar a la de Parkinson, que se llama el principio de Peter: “cada funcionario asciende en su organización hasta alcanzar el máximo nivel de incompetencia”. Claro que no se puede generalizar, no son todos, pero sí es un fenómeno bien conocido, que los funcionarios ascienden hasta llegar a un puesto donde no entienden ni siquiera cuáles son sus objetivos. Los funcionarios que todavía no han alcanzado su nivel óptimo de incompetencia son los únicos que realizan un trabajo efectivo en una empresa. Es una realimentación negativa, que estabiliza a estos funcionarios en sus respectivos puestos donde ya no pueden ascender más. Y eso se nota especialmente en los comités de planeación y de control de calidad. A estos funcionarios, permítanme el chascarrillo, se les podría llamar *bi-reguladores*, porque regulan regular. Controlan lo suficiente para mantener el sistema en la mediocridad. Las personas creativas no suelen terminar en esos puestos porque para ellos es más importante hacer, que controlar lo que hacen los demás. Por otro lado, las personas con muy bajas capacidades no pueden alcanzar esos lugares. De modo que habitualmente es el sitio para los mediocres, en el sentido de que no son buenos, pero no son malos. Están en la mitad. Y puede que ello sea lo correcto socialmente. Colocar allí a gente muy creativa puede estresar demasiado a la organización. El control mediocre se encarga de que la organización funcione, de manera ineficiente, con inercia, con procesos que no están bien diseñados, con mucho desgaste personal, pero logrando que las tareas al final se hagan. Las personas sufren pero la organización simplemente funciona. Si el bosón de Higgs es el que otorga inercia a las partículas, diría de él que es un *bi-regulador*.

Y la conclusión que se puede sacar de todo ello es que la consciencia es lenta, ineficiente, se enreda, tiende a paralizar al organismo.

Para terminar, presento el objetivo de toda esta perorata: pensemos que si la consciencia es burocracia, quizás también debería considerarse que la burocracia es consciencia. En las organizaciones, el afán de interrogarse pidiendo informes, estandarizando comportamientos, especificando funciones, puestos de trabajo y procesos, es la forma que tiene el superorganismo de reflexionar sobre sí mismo, para proponer futuras mejoras. Cumple así con las funciones mínimas de la consciencia. ¿Qué le faltaría? Por un lado, que los informes se traduzcan en acciones para mejoras futuras, cosa que muchas veces no ocurre. Es decir, que se cierren bucles de realimentación. De nada sirve saber por ejemplo que un funcionario no trata bien al público si no se ponen luego mecanismos correctores. Por otro lado, el concepto del *yo*, para lo cual se requiere también que se cierren bucles y que en ellos apa-

rezcan módulos predictivos que intenten anticiparse a los resultados propios y de otros módulos, para que surja un punto fijo.

El surgimiento de la consciencia en el superorganismo lleva acarreada una enorme pérdida de libertad individual, como lo sabemos quiénes estamos ya sufriendo el proceso. Y las etapas iniciales suelen ser muy angustiosas porque los individuos perciben graves errores, típicamente en procesos que van a ser sistematizados con normas hasta volverlos inamovibles aun siendo ineficientes, e incluso existiendo oportunidades de mejora. Porque la consciencia no ve nada de eso, lo único que percibe es que el proceso sistematizado más o menos funciona. Los informes en formatos de papel dicen que todo va bien, aunque la realidad sea otra. A esto se le podría llamar disonancia cognitiva de las organizaciones y quizás conlleve en un futuro a procesos de mejora. Pero el inicio es lento y doloroso.

Y ese es el grave riesgo de la consciencia: su velocidad de adaptación a nuevas situaciones es baja, una vez que ha tomado decisiones respecto a situaciones antiguas. Pensemos por ejemplo que si queremos cambiar un proceso en una empresa, que podría tener un costo⁹⁸ aisladamente de $O(N)$, cuando hay normas, formatos y estándares que lo regulan el costo suele ser del orden $O(N^2)$ debido a las interacciones internas, o quizás más si ello implica cambiar otros procesos de los que depende. Pero lo mismo pasa con la consciencia humana: una vez se ha desarrollado un hábito, una vez que la consciencia ha decidido algo y entrenado al inconsciente para actuar de forma predeterminada, es muy difícil de cambiar. Esto es especialmente palpable en muchas personas mayores, que tienen un conjunto de creencias arraigadas imposibles de modificar aunque les convenzamos con razonamientos de que el mundo a su alrededor ya no es el mismo y se requieren cambios mentales para adaptarse a ello. Decimos que sus conexiones neuronales se han fosilizado. Desgraciadamente pasa incluso con algunos jóvenes, y desde luego, pasa con las organizaciones, lo que siempre me trae a la memoria la portada de uno de los primeros libros sobre sistemas operativos, que muestra a unos dinosaurios enormes atrapados en un pozo de brea.

Para terminar, miremos lo que está ocurriendo ahora mismo. La consciencia en los computadores está más cerca de lo que se piensa. Eso se puede ya palpar en los trabajos de Google (Simonite, 2017) donde para desarrollar los detalles complejos del *software* de inteligencia artificial se usa esa misma inteligencia artificial, lo cual genera un bucle de realimentación positiva, es decir, creativo, que conjeturo que dará lugar a la consciencia en breve.

98 Voy a emplear la notación de complejidad computacional: $O(N)$ significa lineal y $O(N^2)$ significa cuadrático.

Actualmente, a esa inteligencia artificial ya se le hacen preguntas eternas (como “¿cuál es el objetivo de la vida?”) y se están obteniendo respuestas que no son nada filosóficas (como “vivir para siempre”) sino el reflejo del sentir de la humanidad, lo cual es lógico puesto que los datos de entrenamiento proceden de conversaciones humanas. Falta muy poco para que emerja consciencia en las máquinas.

RESUMEN

Hemos dado una explicación a la consciencia, como fenómeno emergente y computacional. El halo de misterio se desvanece. La complejidad que requiere la consciencia es apenas un poco mayor que la de la inteligencia, por lo que es de esperar que todo ser más o menos inteligente tenga consciencia en algún grado, lo cual se ha corroborado en los mamíferos. La inteligencia proporciona capacidad de predicción habitualmente distribuida en módulos. Y la consciencia requiere que esos módulos estén conectados consigo mismos, para que emerja un punto fijo que es lo que llamamos *yo*.

Nos maravillamos de los *qualia* internos, como el color, pero no son más que una forma defectuosa de codificar la información. La información está encriptada porque no ha habido presión evolutiva para mejorar su codificación. En los *qualia* que producen efectos observables no hay esa encriptación y tampoco nos maravillan, porque sí ha habido presión evolutiva para mejorar la codificación.

La consciencia puede entenderse también como la burocracia del cerebro, que no toma decisiones, pero observa, evalúa, genera y pide informes. Al revés, la burocracia en las organizaciones sociales puede entenderse como la emergencia de la consciencia en estos superorganismos. Cuando ello ocurra, y seguramente ya está ocurriendo aunque no lo percibimos, la consciencia a nivel individual humano tenderá a desaparecer, pues entra en conflicto con el nivel superior. Para vivir en sociedad es mejor ser adaptativo, conformista y no ser crítico. Imagina si cada célula de tu cuerpo tuviera consciencia, libertad y una opinión particular sobre cómo hacer las cosas. Por ello es muy posible que a lo largo de generaciones la consciencia humana se vaya apagando. Y nadie se dará cuenta de ello, pues consta únicamente de fenómenos subjetivos. Solo al revisar textos antiguos, escritos antes del siglo XXI, la gente del futuro dirá que hablábamos de ciertos conceptos un tanto esotéricos, que quizás formasen parte del vocabulario de algún ritual mágico.

Por si eso fuera poco, la consciencia en los computadores está a punto de hacer su entrada en escena. ¿Estaremos preparados?

PARA SABER MÁS

Dado que esta presentación sobre la consciencia es un trabajo original mío, no puedo ofrecer recomendaciones de otros libros que hablen más sobre ello. No obstante, ya hay algunos científicos que imaginan que la consciencia es únicamente un patrón matemático, como podemos ver en el siguiente video:

- Max Tegmark (2014). *Consciousness is a mathematical pattern: Max Tegmark at TEDxCambridge 2014*. Recuperado el 16 de octubre de 2017. Disponible en: <https://goo.gl/fY8zdN>

Y desde luego hay excelentes libros que explican muchos más detalles sobre la consciencia, con experimentos de casos reales, como los siguientes:

- Oliver Sacks (1995). *Un antropólogo en Marte*. Barcelona: Editorial Anagrama.

En particular recomiendo este libro fascinante, donde Sacks cuenta muchos casos de autismo, cada uno distinto y de los cuales podemos aprender otra perspectiva de lo que significa “ser humanos”. El autismo no es una enfermedad. Y tampoco tiene una única caracterización sino que hay muchísimas formas de serlo, por lo que hoy se suele hablar del espectro autista o incluso de neurodiversidad. El cerebro tiene muchas arquitecturas alternativas y que la nuestra sea normal mientras que la de los autistas sea la alternativa es bastante discutible. El día que los autistas sean mayoría pensarán que nosotros somos las anomalías. Además, pienso que ser autista puede tener grandes ventajas adaptativas en el mundo que estamos fabricando, donde la consciencia de las máquinas y de los superorganismos (sociedades) está a punto de despegar.

- Daniel Dennett (1995). *La consciencia explicada: una teoría interdisciplinar*. Barcelona: Ediciones Paidós.

Otro libro excelente, donde se habla de la consciencia en todos los aspectos que puede interesar a un ingeniero. Aquí Dennett nos cuenta que no hay ningún homúnculo, que no hay un punto central donde confluyan todas las señales eléctricas de las neuronas y que la consciencia es un fenómeno distribuido por todo el cerebro. Nos explica muchos casos de daños cerebrales y lo que ello implica para la consciencia del sujeto, así como lo que de allí se puede deducir para entender este tema. Mi pequeño aporte personal es anclar los dos fenómenos extraños de la consciencia en el plano computacional. Pero todo lo que había que analizar y decir sobre ella ya lo ha dicho Dennett en este excelente libro.

REFERENCIAS

Libros, artículos y enlaces web

- 123test (2017). *Free IQ test*. Recuperado el 11 de octubre de 2017. Disponible en: <https://goo.gl/Na4jRy>
- Academo (2017). *Wavelength to Colour Relationship*. Recuperado el 5 de febrero de 2017. Disponible en: <https://goo.gl/G6Yfhf>
- Bernardo, Á. (2013). *¿Qué secreto se esconde bajo la ilusión de la mano de goma?* Recuperado el 5 de julio de 2017. Disponible en: <https://goo.gl/3ZVta2>
- Botvinick, M. y Cohen, J. (1998). Rubber hands ‘feel’ touch that eyes see. *Nature* 391(756). DOI: <https://doi.org/10.1038/35784>
- Buller, D. L. (2005). *Adapting mind. Evolutionary Psychology and the Persistent Quest for Human Nature*. Cambridge: The MIT Press.
- Chalmers, D. J. (1995). Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*, 2(3), pp. 200-219.
- Chalmers, D. (2014). *¿Cómo explicamos la consciencia?* TED. Recuperado el 18 de julio de 2017. Disponible en: <https://goo.gl/hfMVSD>
- Crick, F. y Koch, C. (1990). Towards a Neurobiological Theory of Consciousness. *Seminars in the Neurosciences*, 2, pp. 263-275.
- Damasio, A. (2006). *El error de Descartes*. Barcelona: Editorial Crítica.
- DeFelipe, J., Markram, H. y Rockland, K. S. (2012). The Neocortical Column. *Frontiers in Neuroanatomy*, 6(22). DOI: <https://doi.org/10.3389/fnana.2012.00022>
- Dennett, D. (1996). *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Basic Books.
- _____. (2003). *Freedom Evolves*. New York: Penguin Books.
- de Waal, F. (2011). *Comportamiento moral en los animales*. TEDxPeachtree, filmed Nov 2011. Recuperado el 4 de julio de 2017. Disponible en: <https://goo.gl/ML-PZUd>
- Díaz, A. (2015). 35 fantásticas ilusiones ópticas que harán que replantees seriamente tu cordura. Recuperado el 3 de julio de 2017. Disponible en: <https://goo.gl/15UgXH>
- Duro, R. J., Becerra, J. A., Monroy, J. y Calvo, L. (2017). Multilevel Darwinist Brain: Context Nodes in a Network Memory Inspired Long Term Memory. *7th International Work-conference on the Interplay between Natural and Artificial Computation (IWINAC2017)*, A Coruña: Springer Lecture Notes in Computer Science series, 10337, pp. 22-31. DOI: https://doi.org/10.1007/978-3-319-59740-9_3
- Edelman, G. M. (1992). *Bright Air, Brilliant Fire: On the Matter of the Mind*. New York: Basic Books.

- García, A. (2017). Robot's and human's self: a computational perspective. *7th International Work-conference on the Interplay between Natural and Artificial Computation (IWINAC2017)*, A Coruña: Springer Lecture Notes in Computer Science series, 10337, pp. 3-9. DOI: https://doi.org/10.1007/978-3-319-59740-9_1
- _____. (2017b). A Computational Theory of Consciousness: Qualia and the Hard Problem. *Kybernetes*. DOI: <https://doi.org/10.1108/K-10-2017-0387>
- Google (2016). TensorFlow. Recuperado el 29 de noviembre de 2016. Disponible en: <https://goo.gl/DPXXgT>
- Hawkins, J. y Blakeslee, S. (2004). *On Intelligence. How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines*. New York: Henry Holt and Company.
- Hofstadter, D. R. (2007). *I Am a Strange Loop*. New York: Basic Books.
- Hood, B. (2012). *The Self Illusion: How Your Brain Creates You - TAM 2012*. James-RandiFoundation. Recuperado el 15 de julio de 2017. Disponible en: <https://goo.gl/kDYjHD>
- IQ-test.cc (2017). *Test de inteligencia*. Recuperado el 11 de octubre de 2017. Disponible en: <https://goo.gl/WTtxiQ>
- IQ_Research (2017). *IQ TEST*. Recuperado el 11 de octubre de 2017. Disponible en: <https://goo.gl/iJxA4m>
- Kauffman, S. A. (1993). *The Origins of Order*. New York: Oxford University Press.
- Koch, C. (2017). Consciousness in Biological and Artificial Brains. Santa Fé Institute. Recuperado el 18 de julio de 2017. Disponible en: <https://goo.gl/V7Hrb9>
- Kurzweil, R. (2013). *How to Create a Mind: The Secret of Human Thought Revealed*. London: Duckworth Overlook.
- Mensa (2017). *Mensa International*. Recuperado el 11 de octubre de 2017. Disponible en: <https://goo.gl/6Hquesu>
- Mikrojo (2016a). Ilusión Auditiva 'La Paradoja De Shepard'. Recuperado el 3 de julio de 2017. Disponible en: <https://goo.gl/KDPZ7g>
- _____. (2016b). *Ilusión Auditiva 2 'Phantom Words'*. Recuperado el 3 de julio de 2017. Disponible en: <https://goo.gl/YwchQf>
- _____. (2017a). *Ilusión Auditiva 3 'Sonido 3D (Hologonía)'*. Recuperado el 3 de julio de 2017. Disponible en: <https://goo.gl/xZzaXz>
- _____. (2017b). *Ilusión Auditiva 4 'El Efecto McGurk'*. Recuperado el 3 de julio de 2017. Disponible en: <https://goo.gl/wHEeAi>
- Narváez, S., García, A. y Gutiérrez, R. E. (2017). Execution of Written Tasks by a Biologically-Inspired Artificial Brain. *7th International Work-conference on the Interplay between Natural and Artificial Computation (IWINAC2017)*, A Coru-

- ña: Springer Lecture Notes in Computer Science series, 10337, pp. 45-52. DOI: https://doi.org/10.1007/978-3-319-59740-9_5
- Pachniewska, A. (2017). *List of Animals That Have Passed the Mirror Test*. Animal Cognition. Recuperado el 11 de octubre de 2017. Disponible en: <https://goo.gl/sev5Hh>
- Parkinson, C. N. (1955). *Parkinson's Law*. The Economist. November 19th, 1955. Recuperado el 18 de octubre de 2017. Disponible en: <https://goo.gl/WvyTva>
- Pinker, S. (1997). *Como a mente funciona*. São Paulo: Editora Schwarcz.
- Quoc, V. L., Ranzato, M. A., Monga, R. et al. (2012). Building High-level Features Using Large Scale Unsupervised Learning. *The 29th International Conference on Machine Learning – Conference Proceedings*, Edinburgh, pp. 127-137.
- Sacks, O. (2002). *El hombre que confundió a su mujer con un sombrero*. Barcelona: Editorial Anagrama.
- Salgado, R., Prieto, A., Bellas, F. y Duro, R. J. (2017). Motivational Engine for Cognitive Robotics in Non-Static Tasks. *7th International Work-conference on the Interplay between Natural and Artificial Computation (IWINAC2017)*, A Coruña: Springer Lecture Notes in Computer Science series, 10337, pp. 32-42. DOI: https://doi.org/10.1007/978-3-319-59740-9_4
- Sigman, M. (2016). *Your words may predict your future mental health*. TED. Recuperado el 6 de agosto de 2017. Disponible en: <https://goo.gl/bwKB5y>
- Simonite, T. (2017). Why Google's CEO Is Excited About Automating Artificial Intelligence. Recuperado el 18 de julio de 2017. Disponible en: <https://goo.gl/8Bz4NV>
- Sims, K. (1994). Evolved Virtual Creatures, Evolution Simulation. Recuperado el 17 de marzo de 2017. Disponible en: <https://goo.gl/RNDLtC>
- Spinney, L. (2015). The Time Illusion: How your Brain Creates Now. *New Scientist*. Recuperado el 8 de agosto de 2017. Disponible en: <https://goo.gl/mbSjb3>
- Think_Elephants_International (2015). *Mirror Self-Recognition in Asian Elephants!* Recuperado el 11 de octubre de 2017. Disponible en: <https://goo.gl/WYF1EV>
- Tononi, G. y Koch, C. (2015). Consciousness: here, there and everywhere? Royal Society. DOI: <https://doi.org/10.1098/rstb.2014.0167>
- van den Boogaard, E., Treur, J. y Turpijn, M. (2017). A Neurologically Inspired Network Model for Graziano's Attention Schema Theory for Consciousness. *7th International Work-conference on the Interplay between Natural and Artificial Computation (IWINAC2017)*, A Coruña: Springer Lecture Notes in Computer Science series, 10337, pp. 10-21. DOI: https://doi.org/10.1007/978-3-319-59740-9_2

-
- Vsauce (2017). Isolation - Mind Field (Ep 1). Recuperado el 7 de julio de 2017.
Disponible en: <https://goo.gl/KG2Z9T>
- Whitby, B. (1996). *Reflections on Artificial Intelligence*. Exeter, U. K.: Intellect Books.
- Wikipedia (2016a). *Banach Contractive Maps*. Recuperado el 14 de diciembre de 2016. Disponible en: <https://goo.gl/iA5Psi>
- ____ (2016b). *Brouwer fixed-point theorem*. Recuperado el 15 de septiembre de 2016.
Disponible en: <https://goo.gl/1rJ8bA>

Películas y videos

- Lorre, C. y Prady, B. (2007). *The Big Bang Theory*. USA: Warner Bros. Television.
- Roddenberry, G. (1966). *Star Trek*. USA: Desilu Productions & Paramount Television.
- Schepisi, F. (2013). *Words and Pictures*, USA: Roadside Attractions.

CONCLUSIONES

Pues érase una vez un panal de abejas ya terminado ¡por fin! y estas se estaban alistando para ir a hibernar, como los osos polares.

De todos es bien sabido que la vida en colmena es muy mecánica, obedece a una jerarquía de castas y de reglas fijas, cuyo objetivo final es proporcionar la miel a Winnie-Pooh. Ahora mismo se encontraban en la última fase con todo el trabajo hecho y la miel lista para entregar, de modo que estaban quietas, tranquilas y sin nada que hacer.

Se oyó algo así como PLASH - según dirían después las abejas más viejas -y fue que la rama se rompió y se vino abajo con el panal y todo.

La mayoría de las abejas fueron arrastradas hacia abajo sin darse cuenta de que sucediera nada en particular. A las que el accidente las pilló volando, les causó estupor y algunas trataron de seguir su colmena con el objetivo de reintegrarse en la rutina rota por fuerzas desconocidas.

En estas, como el panal estaba dando vueltas mientras caía, una abeja que le seguía de cerca trazó, por casualidad y en su frenesí por la persecución, una figura en el aire parecida a una flor. Algunas otras la vieron y, de entre ellas, algunas vieron el parecido con la flor de la que se alimentaban. Algo se encendió en su pequeño cerebro de abeja al comprender que la flor era y no era a la vez. Se trataba de un símbolo, del primero de un alfabeto que estaba por inventarse - dirían más tarde las abejas semióticas. Ese no fue el único símbolo trazado. Vinieron rápidamente el de panal y el de abeja. Las involucradas en la nueva experiencia, rápidamente aprendieron a trazar en el aire estos dibujos y las que los veían, a su vez aprendían. Luego vinieron las combinaciones de símbolos. Así por ejemplo abeja seguido de flor significaba que yo, la abeja, iba a libar una flor. Mientras que flor seguido de panal significaba que me hubiera gustado llevarme esa flor al panal. De estas y otras muchas combinaciones se pusieron rápida-

mente de acuerdo el pequeño grupo de abejas filólogas que a su vez transmitieron el conocimiento al resto de la comunidad.

Las abejas son extremadamente veloces, todo el mundo lo sabe. Así, mientras el panal caía les dio tiempo a muchas cosas, como por ejemplo, a preguntarse por qué había ocurrido ese BIG-PLASH y que habría habido antes. Y a pensar que con el BIG-PLASH ocurrió algo nuevo que no había sucedido nunca antes: el tiempo. Las más críticas dijeron que el tiempo podía ser una mera consecuencia de la velocidad, mientras que no faltaron opositoras que preferían pensar lo contrario. A todas ellas se las llamó a partir de entonces abejas filósofas, y la cuestión más inquietante que plantearon entonces, y que llegaron a dilucidar demasiado tarde, era si el tiempo, que había tenido un comienzo, tendría también un final.

Generaciones de abejas se sucedieron, creando diversos estilos en el arte (las mil y un maneras de dibujar una flor), en la música (interpretando cada una a su modo el ruido que hizo el BIG-PLASH, en la filosofía (como ya quedó dicho) e incluso en las ingenierías y doctorados.

El sistema de vida tan complejo que se creó, las sutiles relaciones de pensamiento, su dependencia abrumadora respecto a una cultura creada arbitrariamente por ellas mismas, son todos hechos imposibles de entender para alguien que no haya vivido esa experiencia (ser abeja de aquella época).

Creo que nadie alcanzó a escuchar el BIG-CHUF que hizo al estrellarse el panal, cuando finalmente llegó al suelo. Muchas abejas murieron aplastadas, otras de tristeza por el reino perdido, algunas de hambre y las demás deambulan por allí sin poder distinguir lo ocurrido con algún tipo de ensoñación,

El BIG-CHUF

ÁNGEL E. GARCÍA BAÑOS, 2 de julio de 2000

La construcción de este mundo parece debida a una fluctuación del vacío, como dicen los físicos. A un error. Y toda la evolución subsiguiente del universo parece ir encaminada a borrar ese error cuanto antes, sin dejar rastro, en la llamada muerte térmica del universo, que la inteligencia acelera implacablemente. Como consecuencia de ello, y mientras el final llega, experimentamos complejidad.

La teoría actual del nacimiento del universo nos explica que el Big Bang fue exponencial⁹⁹ y las últimas medidas parecen indicar que seguimos así (Goldhaber, 2009). Es lo habitual cuando aparece algo nuevo. Y resulta que hemos experimentado muchos *big bangs* exponenciales si adoptamos una perspectiva mental amplia: la emergencia de la consciencia humana ocurrió en un momento difícil de precisar en un mundo de protohomínidos, incluso

99 Teoría inflacionaria, de Alan Guth en 1981.

quizás bastante después. No había consciencia y poco a poco fue creándose, aumentando de generación en generación hasta llegar al estado actual. Lo mismo puede decirse del lenguaje. No había lenguaje, pero en un proceso de lenta acumulación y de afortunadas coincidencias, se fue creando un lenguaje que terminó siendo lo que vemos y usamos hoy día. Y continúa evolucionando. También ocurrió el *big bang* de la vida, pero que no dejó ni rastro de sus orígenes, pues cualquier estructura autoduplicante inicial fue borrada del mapa por posteriores estructuras más eficientes, en un proceso exponencial de crecimiento, que incluyó también bastantes momentos donde estuvo casi a punto de extinguirse. Los *big bang* no se pueden ver, pues nadie había allí entonces, pero se pueden calcular extrapolando a partir de modelos. Si, como piensa Elon Musk, las formas de vida biológicas somos solo la etapa inicial que da lugar a las formas de vida tecnológicas, es posible que después de un tiempo no quede ni rastro de nosotros, y que para los robots su origen sea un revoltijo de mitos.

Hemos visto que existen varios mecanismos que generan complejidad de forma espontánea, por lo que los mundos, para florecer, no requieren realmente de casi nada. Sin embargo, hay excelentes libros como *Las manchas del guepardo*, de Goodwin, que se empeñan en mostrar a la evolución como un mecanismo menor, casi sin importancia, frente a los procesos matemáticos con dinámicas caóticas o de leyes de potencia. Eso es un error. Cada mecanismo es importante dentro de su ámbito. Mientras la complejidad sea pequeña y no exista la capacidad de autorreplicación, efectivamente imperan los mecanismos del caos. Pero ellos generan una complejidad no adaptativa, no inteligente, incapaz de sobrevivir si el entorno cambia. Solo cuando aparece la autorreplicación y, con ella, la evolución, la complejidad de los sistemas da un gran salto cualitativo y cuantitativo: pueden adaptarse a entornos cambiantes, pueden sobrevivir aunque el mundo se oponga a ellos, y pueden optimizar todo tipo de recursos, logrando de paso que surja la inteligencia y la consciencia. Una vez que aparece la evolución, los procesos caóticos siguen estando allí, de modo que hay una interacción entre ambos que favorece que surja más complejidad. Y de la interacción entre diferentes entes, analizada por la teoría de juegos, aparece nueva complejidad, incluyendo la posibilidad de que los entes se asocien para formar objetos mayores que, a su vez, compitan o cooperen entre sí, formando otros mayores, en una escalera ascendente de complejidad que no pareciera tener fin. Hemos visto una propuesta de algoritmo general para generar complejidad, que se basa en ello. Aplicándolo, la complejidad no tiene límites. A la vez, tampoco tiene garantías y puede derrumbarse en cualquier momento, para volver a crecer a continuación.

En el aspecto de programación se han mostrado los algoritmos evolutivos más importantes, así como una implementación propia del autor de un algoritmo genético básico y un algoritmo generador de complejidad evolutivo. Los algoritmos evolutivos actuales capturan una parte de lo que significa la evolución, pero no todo. Concretamente, los cromosomas nunca se asocian entre sí, para formar cromosomas más complejos. Y la función de aptitud es fija. Ambas cosas limitan las sorpresas a un solo nivel de emergencia. Por así decir, si buscamos sillas, los algoritmos evolutivos nos encontrarán la silla óptima, pero nunca nos ofrecerán un sillón reclinable, una silla con ruedas o un columpio, debido a que la estructura de datos es fija (el cromosoma), lo que impide salirse de un molde prefijado. Además, la función de aptitud codifica rígidamente el objetivo a buscar, sin permitir variantes ni sorpresas en ese sentido.

Mientras que en los aspectos filosóficos se han dado las bases para entender la flexibilidad de la inteligencia. Se ofreció un resumen de las formas de pensar humanas y de los algoritmos existentes hasta ahora. Recordar que inteligencia significa predicción. Y que toda inteligencia es colectiva.

La consciencia es un fenómeno que nos maravilla y que está muy mal entendido. El filósofo Daniel Dennett es quien mejor ha logrado deducir de que se trata: el yo es una ilusión del cerebro, que tiene como objetivo dar un sujeto a la narrativa de todo lo que ocurre. Y el objetivo es también evolutivo: los animales muy simples, como los insectos, posiblemente tienen todas sus neuronas cableadas por los genes, todos sus comportamientos fijos, y no aprenden. El comportamiento de supervivencia y reproducción está cableado, pertenece al *hardware*, y siempre funciona igual, de manera reactiva. Por el contrario, conforme aparece el cerebro y aumenta su complejidad, hay situaciones que obligan a tomar decisiones difíciles. La evolución requiere preservar el individuo, y de ahí el invento del yo en los cerebros superiores. Con él aparece la consciencia y otros epifenómenos como los *qualia* y las sensaciones subjetivas. Todo esto lo hemos aterrizado en este libro en términos computacionales: el yo es un punto fijo que aparece casi inevitablemente cuando tienes un conjunto grande de predictores, conectados unos con otros. Y los *qualia* y las sensaciones subjetivas no son más que fallos de codificación de la información, con funciones que no tienen inversa.

No debemos extrañarnos. Si la inteligencia es un algoritmo, la consciencia debe ser otro algoritmo.

Somos definitivamente robots biológicos, no solo a nivel genético, sino también en lo que tiene que ver con la inteligencia y la consciencia. Y nuestro universo parece determinista. Pero, a pesar de todo, actuamos con libertad, dado que la libertad no es absoluta sino relativa a quienes la miden.

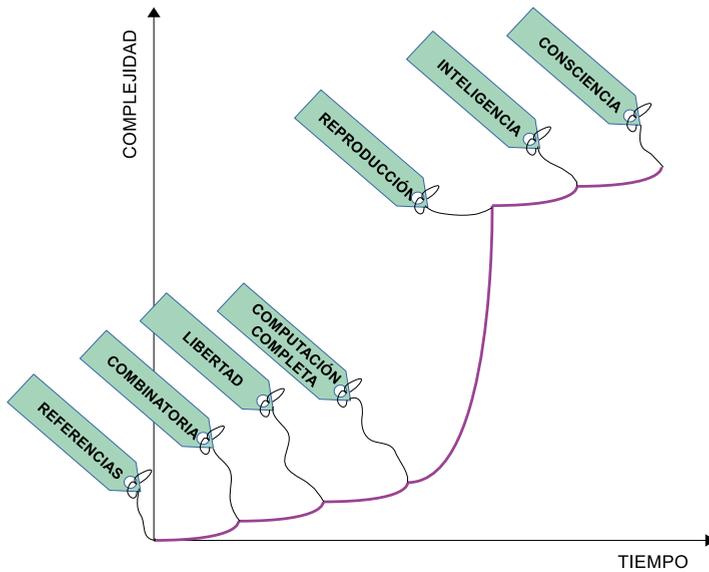


Figura 102. La escalera de la complejidad. Emergencia de propiedades en sistemas complejos

Simultáneamente a todo ello hemos logrado identificar los escalones de complejidad que hay que subir para alcanzar ciertas propiedades (Figura 102). En todos los casos son saltos muy pequeños respecto al nivel anterior, excepto para lograr la reproducción. Son los siguientes:

- **Referencias.** Es el punto de partida. Los objetos deben tener algún nombre u otra forma forma abreviada de referirse unos a otros. Ello no requiere mucha complejidad, sino más bien algún mecanismo físico. De allí es inmediato que aparezcan autorreferencias.
- **Combinatoria.** Si tengo tres objetos $\{A,B,C\}$, pueden aparecer en distintos órdenes $\{ABC, ACB, BAC, BCB, CAB, CBA\}$. Para N objetos hay $N!$ posibilidades. Esto amplía mucho la variedad y complejidad del universo. Es lo que hacen los *quarks* al asociarse para formar *hadrones*, a nivel de partículas elementales. Pero luego, además, puede repetirse con unos niveles dentro de otros, como en grafos donde cada nodo puede ser otro grafo. El universo amplía mucho más su variedad y complejidad con este tipo de hipercombinatoria. Los electrones, protones y neutrones se juntan para formar átomos, que se asocian para formar moléculas y así en muchos niveles.
- **Libertad.** Si hay memoria, algún proceso no lineal y algo de libertad, eso puede producir más libertad. Y si en este nivel no se consigue inmediatamente, se conseguirá cuando aparezca la capacidad de cómputo universal.

- **Capacidad de cómputo universal.** Si hay memoria, comunicación local (que viene dada por referencias de vecindad entre muchos objetos) y cómputo básico (no necesariamente universal, que puede lograrse con transformaciones de símbolos, reacciones químicas, o cambios de estado en general como respuesta al estado anterior y a estímulos de entrada) entonces es bastante probable que aparezcan Máquinas de Turing Universales. La probabilidad de que ello ocurra es del 2%, es decir, muy alta, prácticamente inevitable en un entorno con muchos cambios y muchos objetos, como nuestro universo.
- **Reproducción.** Si un objeto es capaz de sacar copias de sí mismo, entonces sus copias también lo podrán hacer. Y ello automáticamente nos conduce al proceso más importante de generación de complejidad, la evolución, pues adicional a ello, solo se requiere que haya algunos errores al hacer las copias. La probabilidad de emerja la reproducción es muy baja, del orden de $2^{-200000}$, por lo que deben dedicarse más esfuerzos de investigación a tratar de entender las razones. No en vano solo se conocen dos tipos de reproducción en nuestro planeta, la biológica y la de las arcillas (mucho más limitada) y ello es un buen indicador de lo difícil que es la emergencia de esta propiedad. En el siguiente apartado se propondrá una posible solución a este problema. También hay que considerar qué objetos usaremos como punto de partida en la reproducción, pues los animales superiores no comemos tierra, sino aminoácidos, vitaminas y enzimas. La reproducción de robots ¿de qué elementos base partirá? ¿Módulos preensamblados procedentes de fábricas? ¿Minerales que hay que extraer del suelo con los que construir cada pieza? Y si todavía no hay suelo y los átomos vagan en nubes por el espacio ¿habrá que construir primero un planeta donde construir las máquinas que construyan las fábricas que construyan los robots? Quizás, después de todo, se requiera toda una civilización para poder replicar un robot¹⁰⁰.

La reproducción genera nuevos objetos, al igual que la combinatoria. Pero la combinatoria genera objetos completamente distintos, asociando otros objetos preexistentes al azar, sin ejecutar ningún plan, mientras que la reproducción genera objetos muy similares a partir de otro que sirve como patrón. La combinatoria es una búsqueda completamente al azar, mientras que la evolución incluye un bucle de realimentación positiva (cuanto más seres vivos haya, más seres vivos

100 Por analogía con el proverbio africano que dice que se necesita toda una tribu para educar a un niño.

habrá) y otro de realimentación negativa (cuanto más te apartes del óptimo menos te reproduces), que generan un aumento espectacular de complejidad.

- **Inteligencia.** La inteligencia es siempre colectiva. Se requiere un grupo de agentes similares, de modo que para llegar aquí se necesita la reproducción y la evolución. Y poco más. Es necesario que esos agentes interactúen entre sí. La presión evolutiva empuja a los agentes a que predigan el futuro en su afán de sobrevivir (huir de condiciones destructivas y acercarse a situaciones favorables) e incluso para modificarlo en su propio beneficio. La inteligencia emerge inevitablemente si existe previamente el bucle de la evolución. Y la inteligencia es adaptativa, o sea, es mayor cuanto más complejo sea su entorno.
- **Consciencia.** El yo de la consciencia emerge cuando aparece un bucle de realimentación en el conjunto de agentes que realizan predicciones. De nuevo, el salto de complejidad respecto a la inteligencia es mínimo. El *software* actual es capaz de realizar muchas tareas inteligentes, pero todavía no puede reconocerse a sí mismo como un actor dentro del problema que está resolviendo. Pero falta muy poco para ello.

En algunos casos se han podido identificar los bits de complejidad requeridos para alcanzar cada escalón, que son del orden de 8 bits para la capacidad de cómputo universal y 200 000 para lograr la reproducción. En los demás casos (libertad, inteligencia y consciencia) sabemos que se requiere muy poca complejidad, apenas uno o dos bucles adicionales de realimentación, concretamente dos bucles para alcanzar la libertad por medio del caos, un bucle de predicción para la inteligencia y un bucle para la consciencia.

Cabe dentro de lo posible que haya otros escalones de estabilidad que todavía no hemos identificado o no hemos interpretado adecuadamente. Por ejemplo, en mecánica cuántica, las órbitas del electrón en el átomo de hidrógeno son aquellas donde el electrón, visto como onda, interfiere constructivamente consigo mismo. Se puede pensar que todas las partículas y todos los estados de la materia que conocemos son el resultado de una realimentación positiva. Los objetos o procesos que interfieren destructivamente consigo mismos no pueden existir, o tienen un tiempo de vida muy reducido. Esas órbitas del electrón son estables y no hemos hablado de ellas, ni de tantas otras estabilidades.

Y la consciencia no es el punto final. Como vimos en el algoritmo general para producir complejidad, siempre se pueden agrupar entes conscientes para fabricar un nuevo nivel. No hay un límite, pero tampoco hay garantías de crecimiento, pues toda esta construcción es inestable.

Quisiera también resaltar que los temas principales tratados en este libro (complejidad, libertad, vida, inteligencia, consciencia) tienen un factor común: no son fáciles de medir. Y propongo que la forma más razonable de hacerlo es a través de una generalización del test de Turing, donde hay dos agentes compitiendo entre sí para mostrar cuál es más complejo, más libre, más vital, más inteligente o más consciente. Existe un árbitro, que es mejor o igual que ambos agentes en el aspecto a medir. El árbitro interactúa con cada agente por separado, proponiéndole una conversación, o un problema o un obstáculo a ver cómo reacciona. A partir de las respuestas de ambos agentes, el árbitro decide quién es el mejor.

Después hago la observación de que la función de medición de complejidad, libertad, vida, inteligencia o consciencia sufre de *aliasing*, es decir, un árbitro poco complejo (o libre o vital o inteligente o consciente) no debería intentar medir a un agente más sofisticado que él, pues el resultado sería bastante equivocado. Y hago la conjetura que la medición es correcta cuando el árbitro tiene un nivel de sofisticación mayor, igual o ligeramente menor al del agente¹⁰¹. A esta función la llamo mayor o aproximadamente igual (\succeq).

El progreso de las civilizaciones humanas (y en particular de la ciencia) se apoya en esa función, usando una población de árbitros, que pueden ir desde el consejo de ancianos, los pares académicos o los contactos en las redes sociales. Ninguna teoría de ningún tipo ha sido aceptada por la sociedad si su complejidad está demasiado por encima de los árbitros. Debe ser solo un poco mayor¹⁰². Por eso el progreso es lento.

LIBERTAD E INTELIGENCIA

Es un buen momento para reflexionar sobre la relación entre estos dos conceptos y proponer una definición científica de la ética. Recordemos que la libertad nos permite aumentar la inteligencia, y que la inteligencia nos permite aumentar la libertad. Pero que nos permita hacerlo no quiere decir que nos obligue. De hecho, como también comentábamos, la inteligencia también nos permite destruirlo todo de la forma más rápida posible.

Sabemos que el universo aumenta su entropía inexorablemente, de modo que la ética se podría definir de una manera científica como usar la inteligencia para maximizar la libertad, ya que tener mucha libertad significa que

101 Esta conjetura se apoya en observaciones empíricas, pues no hay pruebas en ningún sentido.

102 En la teoría de enseñanza-aprendizaje, Lev Vygotsky propone lo mismo: enseñar cosas nuevas que estén dentro de la zona de desarrollo proximal de los niños (Dixon-Krauss, 1996), es decir, solo un poquito por encima de su nivel actual.

hay muchas posibilidades disponibles, es decir, la entropía se aumentó lo menos posible. Hacemos un esfuerzo, usando nuestra inteligencia y nuestra libertad, para posponer la muerte del universo lo más que se pueda. Estamos hablando ahora de la inteligencia de un individuo y de la entropía del universo completo, que incluye a ese individuo.

De esta manera, asesinar a otra persona (o talar un bosque) no sería ético porque disminuye la libertad de la persona muerta y aumenta la entropía del universo. Y así podría uno llegar a conclusiones matemáticas de cuando conviene matar animales o plantas para comer, si el beneficio en términos de libertad y entropía, compensa el costo. Dicho con otras palabras, la ética consiste en no aumentar gratuitamente la entropía del universo. Evitar comer animales si se pueden comer plantas. Al final, hay que comer, sin duda.

La ética y el aprendizaje (Figura 103) nos sirven para aumentar la libertad y la inteligencia, logrando así que el universo dure mucho. Pero si elegimos otros caminos (en color rojo, en la misma figura), el universo durará menos.

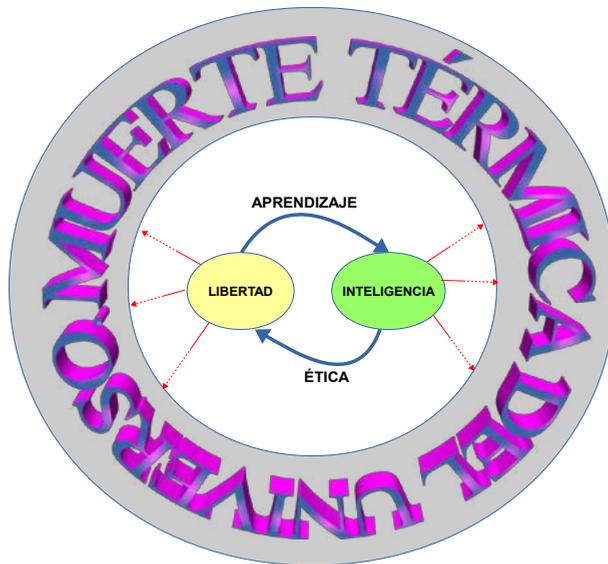


Figura 103. Ética y aprendizaje, en un contexto amplio

OTRA FORMA DE MODELAR LA REPRODUCCIÓN

Este es un tema que quedó pendiente, y quiero retomarlo haciendo una propuesta, ya que partiendo de Máquinas Universales de Turing, la reproducción puede ser más fácil de alcanzar. Tanto en el trabajo de von Neumann sobre reproducción usando autómatas celulares, que vimos en el libro anterior, como en la reproducción de los seres vivos, no se analiza el objeto a

reproducir para sacar una copia de cada una de sus partes. Lo que ocurre es que existe una receta (es decir, un algoritmo) que se lee, se ejecuta y como resultado se fabrica un objeto (Figura 104).

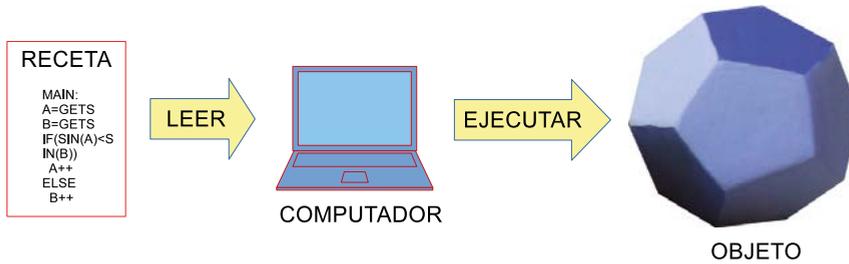


Figura 104. Fabricación de un objeto a partir de su receta

Ahora bien, para que la receta no sea única ni centralizada (lo cual conllevaría el riesgo de que un fallo pueda dar al traste con todo el sistema reproductivo), cada objeto debe de incorporarla en su interior. Hay que pensar en la receta como un programa de computador. Y el computador que ejecuta la receta debe estar disponible ampliamente en el entorno donde viven esos objetos, o bien cada objeto incorporarlo también en su interior (Figura 105).

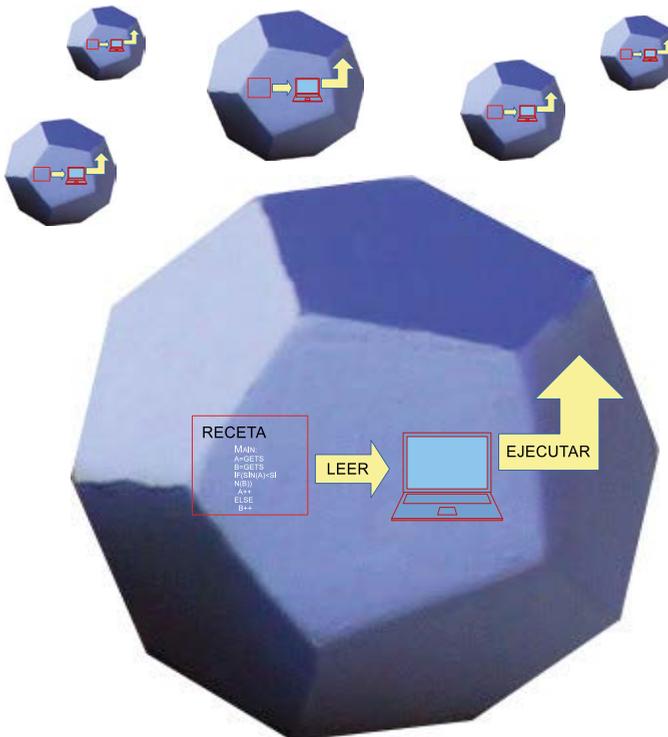


Figura 105. Población de objetos, cada uno con su receta y su computador

Obviamente, la receta debe incorporar también la manera de fabricar ese computador, y seguramente otros detalles de infraestructura básica como la membrana que aísla cada objeto de los demás.

Ya sabemos que las Máquinas Universales de Turing son muy fáciles de obtener por procesos meramente al azar. Lo único que falta por estudiar y entender es como fabricar al azar no la máquina, sino su descripción. Obviamente con ello nos metemos en el típico problema de qué fue antes, si el huevo o la gallina. Para ejecutar la descripción hace falta un computador, y para fabricar el computador hace falta su descripción. A la vez, nada más natural en este tema de la vida artificial que nos aparezca un bucle de este tipo. Entonces ese es el único misterio a resolver: cómo fabricar una Máquina Universal de Turing que contenga una descripción de sí misma. Su complejidad debe ser mucho menor a la de tratar de fabricar un autorreplificador universal. Y para pasar del primero al segundo basta la evolución: por mutaciones de la receta aparecerán objetos nuevos pegados a la Máquina Universal de Turing.

Acaba de descubrirse otro proceso que podría haber originado la vida en la Tierra, y que va en línea con estas ideas (Springsteen, 2018). Inicialmente hay un bucle de autocopiado que se origina con moléculas inorgánicas. Básicamente una reacción autocatalítica, como las que propone Stuart Kauffmann. Pero lo interesante de este trabajo es que muestran cómo cada molécula inorgánica puede ser sustituida por otra orgánica, en pequeños pasos. Y que la reacción química resultante es básica para el ciclo del ácido cítrico, que a su vez es fundamental para la liberación de energía almacenada en las células de todos los organismos aeróbicos.

Sin embargo, podemos desarrollar aún más esta idea. Para ello, hay que recordar que una hipótesis de partida de estos libros es que todo el universo es solo información, y que la Física se puede entender simplemente por medio de procesos computacionales. Entonces, si los objetos de las figuras anteriores son solo información ¿qué es la receta? Pues la misma información, pero comprimida. La Máquina de Turing lee la información y la descomprime, fabricando así el objeto que representa. Pero aún falta algo por entender porque, si bien las MTUs generan información de salida, esa información es muy voluble y se pierde en cuanto la MTU genera más información. La solución a este acertijo es que la materia es la forma de dar persistencia a la información. Los objetos fabricados son materiales, algo concreto, no porque la materia sea dura, por así decir, sino porque consiste en una realimentación positiva que se automantiene.

Pensemos por ejemplo en el *flip-flop* más básico que se puede fabricar con transistores (Figura 106). En él se puede almacenar un bit de informa-

ción gracias a que si el estado Q vale 0, entonces el inversor de arriba lo convierte en 1, imponiendo así este valor al estado complementario \bar{Q} . Y entonces, con Q valiendo 1, el inversor de abajo lo convierte en 0, reforzando este valor al estado original Q . El bit de información se automantiene. Y no hay nada “duro” en esta figura, pues se trata solo de un flujo de información que se resiste a cambiar. La dureza o, más exactamente, la persistencia, tiene que ver con que se requiere energía para cambiar el estado Q a 1, (y su complementario a 0) venciendo la realimentación positiva. La materia cumple la misma función: es la memoria de la información.

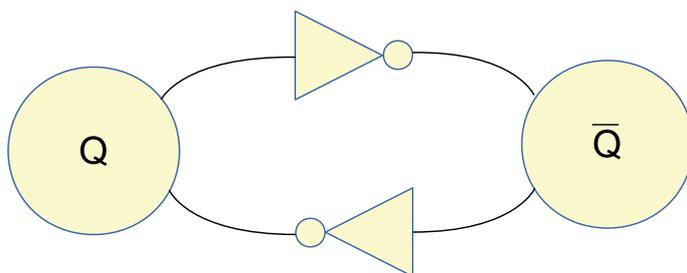


Figura 106. Un flip-flop consiste en dos estados complementarios realimentados por dos inversores

DESPEDIDA

Estamos en una época de gran complejidad. Claro que en cuanto digo eso, me acuerdo del bonito cuento de Borges que al comienzo decía “le tocó, como a todos los hombres, malos tiempos en que vivir”. La verdad es que siempre hemos estado en cambio continuo, de una forma exponencial. Pero las exponenciales al principio son muy lentas, y los cambios que producen a lo largo de la vida de una persona son prácticamente imperceptibles. Nadie se da cuenta. Sin embargo, estamos ahora mismo en un momento donde el crecimiento es muy evidente y la explosión podría terminar mal.

¿Qué ocurre cuando un imperio como el romano crece tanto que las comunicaciones con su frontera se vuelven defectuosas, las órdenes llegan tarde y mal, y nadie tiene información fiable de lo que está pasando en el mundo exterior? El imperio se destruye. ¿Qué ocurre cuando un animal crece tanto que las órdenes del cerebro llegan con demasiado retraso a las extremidades? El animal no puede sobrevivir. ¿Qué ocurre cuando una sociedad crece tanto en complejidad que nadie parece tener el control de lo que sucede? Estamos en momentos de crisis social y la razón es muy obvia: la complejidad de la sociedad ha crecido tanto que no tiene una representación interna de sí misma. Ninguna persona ni subgrupo de personas entien-

de lo que pasa pues, como decíamos antes, la gente acepta mentiras sencillas en vez de verdades complejas. La sociedad ya no tiene capacidad para predecirse a sí misma usando simuladores internos. El futuro es impredecible y está fuera de control. Para evitarlo habría que autolimitar nuestra velocidad de cambio, de modo que siempre tengamos un simulador de nuestra sociedad correctamente sintonizado, funcionando bien y realizando predicciones correctas del futuro.

Cambiando a otro tema, quisiera presentar la reflexión de que la ciencia tradicionalmente es reduccionista, es decir, espera explicar los fenómenos complejos a partir de otros más simples usando relaciones de tipo causa-efecto (Figura 107).

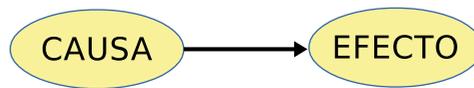


Figura 107. Causa-efecto

Las Ciencias de la Complejidad no desprecian esta vía, pero incorporan otra más potente: la causación circular (Figura 108), que es una clase de realimentación positiva, así como “A” produce “B” y “B” produce “A”. Ello significa que si por alguna razón (bien sea de tipo causa-efecto, bien sea meramente aleatoria) aparece un objeto o proceso que se automantiene (es decir, la causa y el efecto son el mismo objeto, o bien la causa produce como efecto a esa misma causa, quizás directamente o quizás a través de pasos intermedios), esa causación circular es un escalón estable en la escalera de la complejidad. De allí el proceso puede explorar otras transiciones, (bien sean de tipo causa-efecto, o bien aleatorias) para descubrir nuevos escalones de estabilidad, de complejidad mayor. La causación circular tiene una potencia explicativa y funcional mucho más fuerte que la mera causa-efecto. En esta última, si desaparece la causa, desaparece el efecto. En la primera es mucho más difícil desestabilizarla porque se requiere destruir ambos, causa y efecto, a la vez; en caso contrario, cualquiera de ellos regenera al otro.

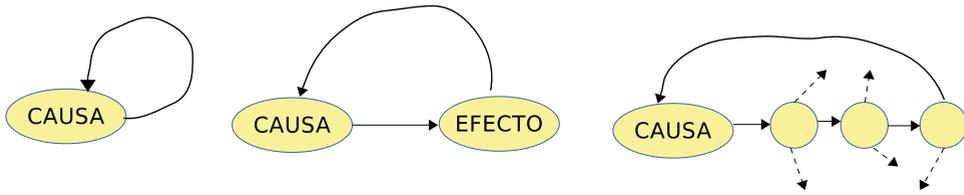


Figura 108. Causación circular

Para terminar, un último objetivo de este libro ha sido que no se vea lo artificial como un concepto peyorativo, ni siquiera radicalmente distinto a lo natural. Actualmente distinguimos entre natural y artificial, no solo en los conceptos aquí abordados (mundos, vida, inteligencia y consciencia) sino también en cosas más cotidianas (comida, medicinas o cultivos). Esa distinción es errónea, por dos razones:

- Lo que ahora llamamos natural y artificial se miden en la misma escala de complejidad. Actualmente usamos lo natural como referencia, únicamente porque tiene más complejidad que lo artificial. Pero la complejidad de lo natural se mantiene constante, al menos en intervalos de tiempo humanamente tratables (100 años), mientras que la complejidad de lo artificial sigue en aumento cada día que pasa y, habitualmente, de forma exponencial. De modo que llegará un momento en que la complejidad de lo artificial supere a la de lo natural. Ray Kurzweil ha estimado que va a ocurrir esa situación aproximadamente en 2045 como consecuencia de la Ley de Moore¹⁰³, y a ese instante lo ha denominado la gran singularidad.
- Lo artificial se define como lo creado por los humanos, lo que no estaba allí por medios naturales. Pero esa definición no se sostiene. Veamos por qué: siguiendo esa definición, las arañas nos parecen naturales. Y las telas de araña también, pues las encuentras en el campo, colgadas de las plantas, sin necesidad de intervención humana. Supongamos ahora que los humanos no existiéramos, y que el animal con inteligencia suficiente para dominar el planeta fueran las arañas. Por analogía, ellas considerarían artificial las telas de araña, porque fueron creadas por ellas, y sin ellas no habría ninguna en el mundo natural. Quizás

103 La Ley empírica de Moore dice que se duplica la capacidad de cómputo de los procesadores cada 2 años. Este es un crecimiento exponencial. Y entonces se puede pronosticar que la capacidad de cómputo de un chip alcanzará a la de toda la humanidad en esta fecha. En fechas cercanas también se estima que se acabará el petróleo y que alcanzarán su apogeo las nanotecnologías y la bioinformática. Algo importante va a ocurrir en esa fecha, aunque no esté claro si va a ser bueno o malo. A ello se le llama la Gran Singularidad.

incluso se quejarían de que hay demasiadas telas de araña en el mundo, y que sería mejor regresar a lo natural. De la misma manera, si los castores fueran la especie dominante considerarían sus represas en las corrientes de agua como algo artificial. Si las hormigas fueran la especie dominante, los hormigueros, la guerra contra otras hormigas, la toma de esclavos entre las hormigas derrotadas y el cultivo de hongos serían consideradas por ellas como algo artificial. Por no hablar de los simios y las herramientas que emplean para comer. En definitiva, los humanos somos miopes al considerar las arañas, los castores, las hormigas y sus productos tecnológicos como algo natural. Todo ser vivo tiene productos tecnológicos. Es lo que Richard Dawkins llama con mucho acierto fenotipos extendidos, en algunos casos demasiado extendidos pues no son resultado directo de la expresión de genes, sino indirecto a través de la construcción de cerebros (como en los homínidos) y la generación de culturas o colectividades (también en homínidos y en hormigas). Pero en cualquier caso no tiene sentido considerar algunos naturales y a otros (únicamente los producidos por los humanos) artificiales.

Y una vez que aceptamos que pueden existir muchos mundos, y que la diferencia principal entre ellos no tiene que ver con lo naturales o artificiales que sean sino con su complejidad, es justo entonces, y exactamente en ese momento, que llegamos al final de este libro.

RESUMEN

En estos dos libros de vida artificial se ha usado un nuevo enfoque para organizar muchos de los conceptos ya existentes. Y, además, en este segundo libro, se han propuesto seis conceptos nuevos:

- Una forma de entender los sistemas complejos basada en medir su complejidad en bits, donde conforme aumenta aparecen nuevas propiedades. Y un algoritmo general evolutivo para crear sistemas complejos.
- Una definición positivista de libertad, como la capacidad de evitar ser predicho.
- Una definición positivista de inteligencia, como la capacidad de predecir. Y una clasificación nueva de las formas de razonar.
- Un abordaje computacional a la consciencia, donde se resuelven los problemas del yo, de los *qualia* y de la experiencia subjetiva (*hard pro-*

blem), que permitirán fabricar consciencia artificial en robots y en organizaciones humanas.

- Una forma de definir la reproducción que es más alcanzable, pues requiere menos complejidad.
- Una definición científica de la ética.

REFERENCIAS

Libros, artículos y enlaces web

Borges, J. L. (2007). *Nueva refutación del tiempo. Otras inquisiciones*. España: Destino.

Dixon-Krauss, L. (1996). *Vygotsky in the classroom*. New York: Longman Publishers.

Goldhaber, G. (2009). The Acceleration of the Expansion of the Universe: A Brief Early History of the Supernova Cosmology Project (SCP). *AIP Conference Proceedings*, 1166(53). arXiv:0907.3526. DOI: <https://doi.org/10.1063/1.3232196>

Springsteen, G., Reddy, J., Nelson, J., Joel, C. y Krishnamurthy, R. (2018). Linked cycles of oxidative decarboxylation of glyoxylate as protometabolic analogs of the citric acid cycle. *Nature Communications*. DOI: <https://doi.org/10.1038/s41467-017-02591-0> Recuperado el 18 de enero de 2018. Disponible en: <https://goo.gl/EUdNEZ>

Tesis y trabajos de grado en EVALAB

En EVALAB por distintas circunstancias se han trabajado otros temas de *software* que nada tienen que ver con la vida artificial. Aquí los presento, principalmente como un reconocimiento a estos estudiantes que también participaron en el grupo aportando ideas, ayudas, conocimiento y buen ambiente.

Arango, L. H. (2005). *Sistema de recolección de información para equipos de tecnología informática*. Cali: Universidad del Valle.

Becerra, A. (2003). *CORBA para Mozart*. Cali: Universidad del Valle.

Bolaños, J. J. (2008). *Aplicación web para el manejo de información académica del colegio técnico Comfacauca School Manager*. Cali: Universidad del Valle.

Caicedo, S. H. (2009). *Desarrollo de una aplicación software identificador de factores protectores y de riesgo asociados al VIH - Sida y ETS*. Cali: Universidad del Valle.

Campo, H. A. (2004). *Estudio, análisis y validación del estándar CORBA - El servicio de seguridad*. Cali: Universidad del Valle.

Castañó, L. (2000). *Administración de transacciones en CORBA*. Cali: Universidad del Valle.

- Chamat, O. A. (2014). *Desarrollo de una aplicación Web para automatización de entrega y corrección de trabajos de programación*. Cali: Universidad del Valle.
- Gálvez, L. E. (2014). *Sistema centralizado de autenticación y autorización "Single Sign On"*. Cali: Universidad del Valle.
- García, P. A. (2013). *Sistema adaptativo de enseñanza del idioma sobre la plataforma Moodle*. Cali: Universidad del Valle.
- González, R. E. (2009). *Diseño de un algoritmo criptográfico Karn-SHA2-Feistel, mediante combinación de primitivas*. Cali: Universidad del Valle.
- Gutiérrez, J. E. (1999). *Diseño y construcción de un sistema operativo con fines educativos*. Cali: Universidad del Valle.
- Huguett, J. E. (2008). *Comparación entre lenguaje PHP y Ruby e implementación de una aplicación de comercio electrónico en Ruby on Rails*. Cali: Universidad del Valle.
- Marín, C. D. (2014). *Creación de novedad del sistema de información de recursos humanos (SIRH) correspondiente a la nivelación salarial de docentes y trabajadores oficiales*. Cali: Universidad del Valle.
- Martínez, M. del P. y Rojas, J. (2005). *Aplicación web para el manejo de la información académica de la corporación G.I.Q.* Cali: Universidad del Valle.
- Medina, J. J. (2006). *Creación de una especificación para un modelo de programación distribuida orientada a eventos*. Cali: Universidad del Valle.
- Melgarejo, M. M. (2005). *Informe técnico del diseño e implementación del sistema de información S.A.M. (seguimiento, administración y manejo)*. Cali: Universidad del Valle.
- Ochoa, A. A. (2009). *Servicio de directorio con listas de correo para la EISC en la Web (SDLCEW)*. Cali: Universidad del Valle.
- Payán, F. A. (2012). *Sistema integral de calificación automatizada como alternativa para optimización y gestión de evaluación de examen tipo test en la Universidad del Valle*. Cali: Universidad del Valle.
- Reyes, J. M. (2015). *Yuto - modelo de adaptatividad de ayudas para entornos e-learning utilizando Agentes Software*. [Tesis Maestría]. Cali: Universidad del Valle.
- Salazar, J. P. (2006). *Aplicación web para la administración y envío masivo de mensajería a telefonía móvil*. Cali: Universidad del Valle.

SOLUCIÓN A PROBLEMAS DE INGENIO

Lástima que hayas tenido que llegar hasta aquí. Pero no te preocupes, que no se lo diremos a nadie.

LAS 12 MONEDAS

Recuerdo que este problema me lo contó el profesor Alfons Crespo, de la Universidad Politécnica de Valencia. Lo primero que uno intuitivamente piensa es en dividir las monedas en dos grupos de seis, y poner un grupo en cada platillo. Pero, a poco que reflexiones, te darás cuenta que esa medida prácticamente no da información. Uno de los platillos va a pesar más que el otro, inevitablemente. O sea, que no hemos averiguado nada que no supiéramos desde el principio.

La solución está en dividir las monedas en tres grupos de cuatro. Un grupo se deja sobre la mesa y los otros dos se llevan a los respectivos platillos de la balanza. Si, por ejemplo, los dos platillos pesan igual, entonces la moneda falsa está sobre la mesa y nos podemos centrar en averiguar cuál es de entre cuatro posibles. Este tipo de medida sí nos da mucha información.

Claro que también puede ocurrir que la balanza se incline hacia un lado. Eso, de todos modos, nos permite descartar las cuatro monedas que quedan sobre la mesa.

Como hay varios casos que tratar, he diseñado este diagrama (Figura 109) para cubrirlos todos. En el lado izquierdo aparecen las doce monedas que se pueden situar en tres sitios, donde “M” (o blanco) significa la mesa, “I” significa el platillo izquierdo y “D” el derecho.

Cada una de las tres pesadas (en las siguientes tres columnas) pueden arrojar tres resultados: “I>D” significa que el platillo de la izquierda pesa más que el de la derecha; “D>I” es la situación inversa; y “I=D” es cuando ambos pesan igual. En la siguiente columna, en amarillo, está el resultado. Concretamente, el número de la moneda falsa y si pesa más (indicado con “+”) o menos (“-”) que las demás. Las dos siguientes columnas ofrecen información adicional sobre el proceso.

El total de bits de información que nos arroja este sistema es $\log_2(2^3 \cdot 12) = 4.58496$ bits. Y en cada pesada se obtiene aproximadamente un tercio de esta cantidad, es decir, 1.52832 bits. O sea, casi 1 *trit* (definiendo *trit* como la información que se obtiene al hacer una pregunta que tiene tres respuestas equiprobables). No se alcanza el *trit* porque hay tres casos imposibles, como se ve en la figura.

Es decir, como hay tres preguntas (las tres pesadas) y cada una puede dar tres respuestas, entonces hay $3^3 = 27$ resultados posibles. Pero como solo hay 24 casos (cada una de las 12 monedas puede pesar más o menos que las demás), entonces hay $27 - 24 = 3$ casos imposibles.

Monedas												CONOCIMIENTO antes de la jugada					
1	2	3	4	5	6	7	8	9	10	11	12	NO DIFINIDO	incertidumbre				
I	I	I	D	D	D	D	M	M	M	M		NO DIFINIDO	no sabemos nada	24/24			
I	I	D	D	I	M	M							I > D	1+ or 2+ or 3+ or 4+ or 5+ or 6+ or 7+ or 8-	8/24		
I	I	D	D	I	M	M							I = D	9+ or 9- or 10+ or 10- or 11+ or 11- or 12+ or 12-	8/24		
I	I	D	D	I	M	M							D > I	1- or 2- or 3- or 4- or 5+ or 6+ or 7+ or 8+	8/24		
I	D												I > D	(1+ or 2+ or 3- or 4- or 5- or 6+) and B => 1+ or 2+ or 5-	3/24		
I	D												I > D	(7+ or 7- or 8+ or 8-) and B => 7- or 8-	2/24		
I	D												I > D	(1- or 2- or 3+ or 4+ or 5+ or 6+) and B => 3+ or 4+ or 6-	3/24		
I	D												I = D	(8+ or 9+ or 10- or 11-) and C => 9+ or 10- or 11-	3/24		
I	D												I = D	(12+ or 12-) and C => 12+ or 12-	2/24		
I	D												I = D	(8- or 9- or 10+ or 11+) and C => 9- or 10+ or 11+	3/24		
I	D												D > I	(1+ or 2+ or 3- or 4- or 5- or 6+) and D => 3- or 4- or 6+	3/24		
I	D												D > I	(7+ or 7- or 8+ or 8-) and D => 7+ or 8+	2/24		
I	D												D > I	(1- or 2- or 3+ or 4+ or 5+ or 6+) and D => 1- or 2- or 5+	3/24		
I	>	D											I > D	1+	1/24		
I	>	D											I > D	I = D	5-	1/24	
I	>	D											I > D	D > I	2+	1/24	
I	>	D											I > D	I = D	I > D	impossible	
I	>	D											I > D	I = D	I = D	7-	1/24
I	>	D											I > D	I = D	D > I	8-	1/24
I	>	D											I > D	D > I	I > D	3+	1/24
I	>	D											I > D	D > I	I = D	6-	1/24
I	>	D											I > D	D > I	D > I	4+	1/24
I	=	D											I = D	I > D	I > D	11-	1/24
I	=	D											I = D	I > D	I = D	9+	1/24
I	=	D										I = D	I > D	D > I	10-	1/24	
I	=	D										I = D	I = D	I > D	12-	1/24	
I	=	D										I = D	I = D	I = D	impossible		
I	=	D										I = D	I = D	D > I	12+	1/24	
I	=	D										I = D	D > I	I > D	10+	1/24	
I	=	D										I = D	D > I	I = D	9-	1/24	
I	=	D										I = D	D > I	D > I	11+	1/24	
D	>	I										D > I	I > D	I > D	4-	1/24	
D	>	I										D > I	I > D	I = D	6+	1/24	
D	>	I										D > I	I > D	D > I	3-	1/24	
D	>	I										D > I	I = D	I > D	8+	1/24	
D	>	I										D > I	I = D	I = D	7+	1/24	
D	>	I										D > I	I = D	D > I	impossible		
D	>	I										D > I	D > I	I > D	1-	1/24	
D	>	I										D > I	D > I	I = D	5+	1/24	
D	>	I										D > I	D > I	D > I	2-	1/24	

Figura 109. Solución al problema de las 12 monedas

¿CUÁL ES EL SISTEMA MÁS COMPLEJO?

Respuesta: (c). Como veremos a lo largo del capítulo, hay muchas formas de medir la complejidad, pero lo intuitivo aquí es tratar de describir la sala de cómputo con palabras. Podemos nombrar los computadores de izquierda a derecha y de arriba a abajo así: A, B, C, D, E y F. Entonces las más simples son la sala (a) y la sala (d), la siguiente la (b) y la siguiente la (c). Podría parecer que la más compleja es la sala (d) porque tiene más conexiones, pero ello no es así.

- (a) Son 6 computadores todos desconectados.
- (b) Son 6 computadores. Está conectado A con B.
- (c) Son 6 computadores. Están conectados A con C; A con E; B con C; B con D; B con E; B con F; C con D; C con E; D con E; y D con F.
- (d) Son 6 computadores conectados todos con todos.

ALGORITMO DESCRIPTOR

La primera secuencia (1, 3, 5, 7, 9, 11, 13, 15, 17, 19) es trivial, son los primeros 10 números impares.

La segunda secuencia (1, 4, 1, 5, 9, 2, 6, 5, 3, 5) es un poco más difícil: son las 10 primeras cifras decimales del número $\pi = 3,1415926535\dots$

La tercera secuencia (0, 5, 4, 2, 9, 8, 6, 7, 3, 1) es bastante difícil, especialmente para ingenieros, científicos o matemáticos: son los números del 0 al 9 escritos lexicográficamente y puestos en orden alfabético. Es decir:

Cero
Cinco
Cuatro
Dos
Nueve
Ocho
Seis
Siete
Tres
Uno

Quizás aquí nos damos cuenta que hay demasiados algoritmos a considerar, para intentar expresar una secuencia de la manera más corta posible. Encontrar el más corto depende de la inteligencia del programador, y no hay una forma automática de lograrlo, ni de garantizar que no haya una descripción todavía más corta.

EL BAR “EL FAROL”

Todos los participantes son iguales, en el sentido de que están sometidos a las mismas reglas del juego y usan el mismo algoritmo. Lo interesante aquí es que si el algoritmo es determinista, entonces está garantizado que no encuentran la solución, incluso obtienen el peor resultado posible. De modo que es necesario emplear algoritmos aleatorios. El determinismo suele servir para solucionar problemas sencillos. Mientras que para solucionar problemas complejos se requiere una inteligencia distinta, que incorpore aleatoriedad, es decir, que tenga libertad de decisión. Y ese es el tema que estamos hablando en el capítulo “Libertad” y al que volveremos en más en detalle cuando hablemos de **estrategias mixtas**, en el capítulo “Teoría de Juegos”.

En Arthur (1994) se plantea este juego y se propone una forma computacional de resolverlo, jugando muchas veces y usando un conjunto diverso de predictores, eligiendo de entre ellos el que mejor se vaya aproximando a una buena solución. El resultado, que podría haberse obtenido también usando algoritmos evolutivos, es muy simple: cada participante debe lanzar al aire una moneda sesgada con un 60% de probabilidad de que salga cara. Si sale cara debe ir al bar y, en caso, contrario, debe quedarse en casa. Eso asegura que, en promedio, cada viernes, el 60% de la población vaya al bar.

TEOREMA DE LA AMISTAD

La solución es $N=6$ personas. Para demostrarlo primero veamos contraejemplos para $N=3$, $N=4$ y $N=5$ en la figura 110 donde los nodos representan las personas y los arcos representan la relación: {rojo=no se conocen, verde=se conocen}.

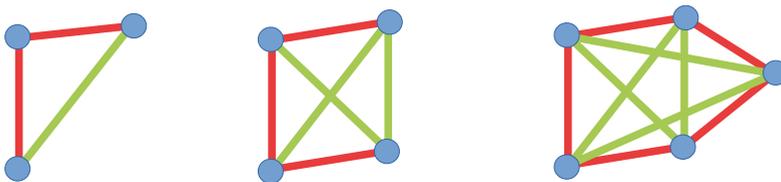


Figura 110. Fiestas con 3, 4 y 5 personas, que no cumplen el teorema de la amistad

Para $N=6$ elijamos un nodo de cualquiera que vamos a llamar A. De allí salen 5 arcos a los demás nodos. Al menos 3 arcos deben de ser del mismo color, que vamos a suponer verde sin pérdida de generalidad. Esos 3 arcos terminan en 3 nodos que vamos a llamar B, C y D. Si alguno de los arcos BC, BD o CD es verde (Figura 111-a), ya tenemos un triángulo verde, lo que

significa que hay 3 personas que se conocen entre sí. Pero si ninguno de los arcos BC, BD y CD es verde, entonces los tres arcos son rojos (Figura 111-b), lo que significa que B, C y D no se conocen entre sí. QED.

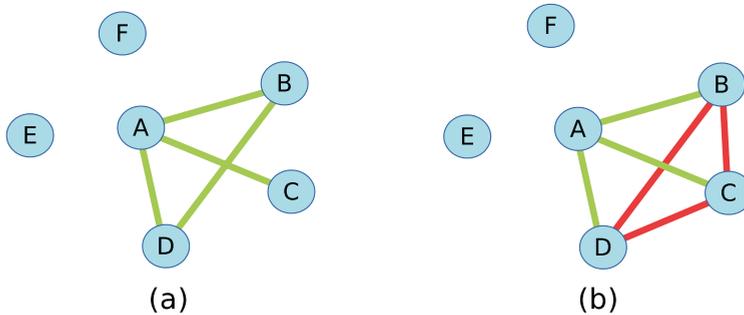


Figura 111. Dos casos en los arcos BC, CD y BD: (a) al menos hay un arco verde; (b) no hay ningún arco verde

GALLETAS PARA TODOS

El valor mínimo es $N=44$. La demostración se hace por inducción:

Primero, supongamos que ya sabemos cuál es ese valor N . Entonces, si nos llega un pedido de $N+1$ galletas lo podremos atender eliminando una caja de 20 y añadiendo cajas de 9, 6 y 6 galletas, así:

$$N+1 = N+9+6+6-20$$

La notación aquí es así:

- En el lado izquierdo de la igualdad, “+X” significa sumar X galletas.
- En el lado derecho de la igualdad, “+X” significa añadir una caja de X galletas, lo cual también es una suma de galletas. Y “-X” significa eliminar una caja de X galletas, lo cual también es una resta de galletas.

De la misma manera para pedidos de cantidades sucesivas:

$$N+2 = N+20-9-9$$

$$N+3 = N+9-6$$

$$N+4 = N+6+6+6+6-20$$

$$N+5 = N+20-6-9$$

$$N+6 = N+6$$

En este momento no es necesario continuar hasta $N+\infty$ porque lo que nos dice la última fórmula es que si sabemos empacar N galletas, para lograr empacar $N+6$ simplemente se requiere añadir una caja de 6. Por tanto, a partir de aquí es repetir lo mismo, sustituyendo N por $N+6$. Por ejemplo:

$$N+7 = (N+6)+1 = (N+6)+9+6+6-20$$

Para terminar, hay que determinar qué valor de N permite hacer esto. Debe ser un N al cual se le puedan quitar cajas de 20, 9, 9 y 6, lo cual suma $20+9+9+6=44$.

Por tanto, el pedido a partir del cual se pueden atender todos los demás es $N=44$. QED.

ACERTIJO MU

Para llegar a la solución hay que buscar algún método por el que podamos obtener el resultado en un tiempo finito. Es lo que se llama un proceso de decisión o, en términos computacionales, un algoritmo. Si pensamos en programación, tendríamos que construir un programa que aplique todas las reglas posibles a MI y a todas las verdades que vaya generando, hasta lograr convertirla en MU. Pero si lo hacemos, nos daremos cuenta de que el programa no para. Pasa el tiempo días, años, milenios, y el programa no se detiene. No encuentra la solución. Entonces uno se pregunta si MU es falso, o bien si MU es verdadero pero su demostración es muy compleja.

Puedes intentar hacerlo con papel y lápiz. O, mejor aún, escribiendo tu propio programa. Pero pasado un cierto tiempo verás que no logras deducir MU de MI. ¿Eso es debido a que no has tenido suficiente paciencia y debes explorar más?

La solución a este problema no se logra aplicando las reglas. Hay que analizar las reglas y saltar afuera del sistema, creando otro distinto. Veamos como:

La “M” es irrelevante. Nunca se quita ni se pone ni se modifica.

El número de “I” solo lo podemos aumentar usando la regla 2, duplicando la cantidad.

El número de “I” solo lo podemos disminuir por múltiplos de 3 (usando la regla 3). Por tanto, para eliminar todas las “I” tenemos que conseguir que aparezca un múltiplo de 3 veces.

Un número que no es divisible por 3, sigue sin serlo después de multiplicarlo por 2 (si usamos la regla 2).

Un número que no es divisible por 3, sigue sin serlo después de restarle 3 varias veces (si usamos la regla 3).

Además, 1 (el axioma inicial tiene 1 “I”) no es divisible por 3.

Resumiendo, para que el problema tuviera solución se necesita alcanzar un número de “I” que sea potencia de 2 y divisible por 3, lo cual es imposible.

O sea, no existe un n tal que $n=2^a$ y $n=3^b$ siendo n, a, b números naturales

De ahí se deduce que MU no es alcanzable a partir de MI, aplicando las reglas. Es decir, MU no es un teorema del sistema, o sea, no es una verdad

del sistema. Sin embargo, no podemos deducirlo usando únicamente el sistema de reglas inicial.

Se sospecha que puede pasar lo mismo con algunas de las conjeturas de las matemáticas. Por ejemplo, la conjetura de Goldbach dice que todo número par mayor que 2 es suma de dos primos. Podemos construir un programa que lo verifique, por ejemplo:

$$4 = 2 + 2$$

$$6 = 3 + 3$$

$$8 = 3 + 5$$

$$10 = 5 + 5$$

$$12 = 5 + 7$$

$$14 = 7 + 7$$

$$16 = 5 + 11$$

Etcétera.

Ec. 39

Pero hasta ahora no se ha logrado ni un contraejemplo ni una demostración para la conjetura de Goldbach.

DOS AUTOREFERENCIAS DIVERTIDAS

Las dos autorreferencias fueron tomadas del libro *Gödel, Escher, Bach: un grácil e infinito bucle*, de Douglas Hofstadter (1979).

Para que diga la verdad, la frase debe de ser: “*Esta frase contiene cuarenta y siete caracteres*”. Por cierto que no hace falta tener un PhD. en Matemáticas para inventar estas cosas. Entre La Pintada y el Alto de Minas, en Colombia, hay un bar de carretera que se llamaba “*Estadero veintiséis letras*”.

La siguiente frase quisiera decir de sí misma “*Esta frase no tiene verbo*”, pero en ese caso estaría mintiendo. Para decir la verdad no le queda otra alternativa que suprimir su propio verbo.

EL BARBERO DE SEVILLA

Es una versión más elaborada de la afirmación de Epiménides el cretense, por lo que no se puede saber la respuesta. Si suponemos que sí, el enunciado nos dice que no. Y si suponemos que no, el enunciado nos dice que sí. En cualquier caso llegamos a una contradicción, por lo que la pregunta del enunciado es indecidible, de manera similar a cómo no se estabiliza la señal digital de una puerta *NOT* cuando se conecta su entrada a su salida.

PARADOJA DE RICHARD

La paradoja de Richard en la variante para números enteros está mal entendida. Veamos por qué:

Se listan todas las propiedades de los números enteros y se ordenan lexicográficamente. Después se calcula el conjunto de números que cumplen con esa propiedad. Por último, se mira si el número de orden está dentro de ese conjunto (y entonces se le llama no-Richardiano) o no (y entonces se le llama Richardiano).

Tabla 2. Lista de números Richardianos y no-Richardianos

Número	Propiedad	Conjunto de números que cumplen con la propiedad	¿Richardiano?
1		C_1	
2		C_2	
...			
R	No cumplir con la propiedad que representa	$C_R = ?$?
...			
17	No ser divisible por ningún otro entero excepto por la unidad y por él mismo	Los primos: $C_{17} = \{2, 3, 5, 7, 11, 13, \dots\}$	NO
...			
22	Ser producto de un entero por sí mismo	Los cuadrados: $C_{22} = \{1, 4, 9, 16, 25, 36, \dots\}$	SÍ
...			

El problema radica en que los conjuntos de números que cumplen con la propiedad enunciada son estáticos, excepto para la propiedad R, en que depende de R.

Es decir, no existe la lista de números no-Richardianos, porque depende de la propiedad que representen. Por ejemplo, para la propiedad 17 del ejemplo anterior, la lista de números no-Richardianos es $\{2, 3, 5, 7, 11, 13, \dots\}$. Y para la propiedad 22 es $\{1, 4, 9, 16, 25, 36, \dots\}$

Es decir, no existe un conjunto de números estático para la propiedad R, de modo que esa propiedad no está bien definida.

Por otro lado, se encuentra frecuentemente una argumentación errónea (RP_Wiki, 2017; Hofstadter, 1979), que dice más o menos así:

Supongamos que R es Richardiano. Entonces no debe cumplir con la propiedad que representa. Por tanto R es no-Richardiano.

La tercera afirmación no sé de dónde la sacan. Es coherente que R sea Richardiano y no cumpla con su propiedad.

Y luego siguen:

Supongamos que R es no-Richardiano. Entonces debe cumplir con la propiedad que representa. Por tanto R es Richardiano.

De nuevo, la tercera afirmación no sé de dónde la sacan. Es coherente que R sea no-Richardiano y cumpla con su propiedad.

La indecibilidad no viene por contradicción, sino porque ambas proposiciones, la directa y la negada, son válidas. Parece que el lenguaje, cuando tiene frases con varias negaciones, nos confunde.

Observemos también que C_R (de la tabla 2) no es un conjunto simple, sino un conjunto de conjuntos, cuyo elemento R-ésimo es también un conjunto de conjuntos, cuyo elemento R-ésimo también lo es, y así infinitas veces. O sea, que C_R no es constructible, pues requiere una recursión infinita.

Y justo ese elemento R-ésimo, del elemento R-ésimo, del elemento R-ésimo, repetido infinitas veces, es el que hay que mirar para ver si R está allí.

Ambas posibilidades son válidas: podemos decidir que R está allí o podemos decidir que no está. Sin embargo, dado que no hay un elemento final, pues es una recursión infinita, y que los elementos anteriores al final son todos conjuntos (es decir, no son números), lo más razonable es decidir que R no está allí. Por tanto, R es Richardiano.

LO QUE HENRY DECIDIÓ

Si todos los coches dañados tenían el timón perfecto, eso significa que se está gastando demasiados recursos en el timón. Henry Ford puede ahorrar dinero haciendo timones no tan perfectos para que fallen más o menos a la vez que el resto de las piezas del coche. ¿Y cuánto tiempo es eso? Se puede calcular en función de los accidentes. Por ejemplo, si resulta que en promedio un coche tiene un accidente grave a los 20 años, es inútil hacer piezas que duren mucho más tiempo.

Con los seres vivos, incluyendo los humanos, ha pasado lo mismo. Los fallos en las piezas, o sea, las enfermedades de la vejez tales como cáncer, problemas cardíacos, fallos cerebrales, etc., habitualmente aparecen a partir de los 60 años porque en el ambiente tradicional en el que hemos vivido más tiempo y donde la evolución ha modelado nuestros genes, era muy probable morir de cualquier infección por alguna herida o alguna caída o alguna mordida por un animal antes de llegar a los 35 años de edad. La evolución no tenía ningún incentivo en fabricar corazones que durasen 500 años si su propietario, en promedio, iba a morir a los 35.

Ahora, gracias a que en los últimos cientos de años hemos pasado a vivir en grandes sociedades que tienen la medicina, la seguridad y la higiene como subproductos, las infecciones ocurren menos y no suelen ser mor-

tales, las mordidas de león son casi imposibles y si te caes rompiéndote un hueso hay hospitales donde te operan y te salvan la vida. Como resultado, la edad promedio humana se acerca a los 90 años en muchas partes del mundo. Y los genes, en estos pocos cientos de años, no han tenido tiempo de darse cuenta de ello, y no han logrado adaptarse a la nueva situación. Como resultado, vemos una proliferación de enfermedades degenerativas que antes eran anecdóticas.

Además hay otro problema: nuestra etapa reproductora va aproximadamente desde los 15 años de edad hasta los 40. El otro objetivo de la evolución es la reproducción, de modo que tampoco hay mucha presión evolutiva por lograr personas que duren más allá de los 40 años si no se van a reproducir.

Todos estos problemas no deben desanimarnos. Actualmente hay muchos programas de investigación puntuales, dirigidos a prolongar la edad de los humanos por medio de mejores estilos de vida, nuevas medicinas y también de genética. Se trabaja en:

- Reparar genes defectuosos para evitar que produzcan enfermedades.
- Identificar los genes que determinan mayor longevidad para poder implantarlos en humanos.
- Identificar proteínas y enzimas en la sangre o en otros tejidos que promueven la longevidad.
- Medicinas, vitaminas y antioxidantes para preservar las funcionalidades de las células y evitar daños en ellas.
- Comida y estilos de vida más sanos. El estrés y la contaminación del aire y del agua producen envejecimiento.
- Nuevas prácticas médicas. El médico conocerá tus genes y te aplicará tratamientos de mantenimiento personalizados para tu cuerpo.
- Nanorobots que puedan reparar daños internos.

Pero hay además un programa de investigación integral para tratar la vejez de forma similar a como lo hacemos con un automóvil, es decir, con mantenimiento básico y reparación de los problemas que surjan. Lo dirige el gerontólogo biomédico Audbrey de Grey (2017) que ha identificado y propuesto solución para siete grandes grupos de enfermedades:

- Pérdida de células por lesiones, traumas, necrosis, etc. que produce enfermedades como el Parkinson y la diabetes. El tratamiento consistirá en inyectar células madre que regeneren los tejidos dañados.
- Resistencia a la muerte celular, que suele producir sobre todo obesidad. El tratamiento será enviar fármacos o virus que eliminen selectivamente solo esas células.

- Reproducción celular sin control, producida por daños en el ADN y que típicamente se identifica como cáncer. La respuesta a este problema es algo más complicada e invasiva: consiste en neutralizar los telómeros del ADN de todas las células del cuerpo, para impedir que se reproduzcan sin control, y resembrar todo el cuerpo con células madre cada 10 años, para facilitar la sustitución de células envejecidas o dañadas.
- Aumento de la basura intracelular, que produce arterioesclerosis y degeneración macular seca, entre otras enfermedades. Para evitarlo, de Grey propone diseñar e introducir nuevas enzimas en las células, capaces de descomponer esa basura.
- Aumento de la basura extracelular, que produce Alzheimer, diabetes tipo II, etc. El tratamiento consistiría en desarrollar vacunas para que el propio sistema inmune descomponga esa basura.
- Rigidez de los tejidos debido a reacciones de glicación producidas por los azúcares. El fenómeno adverso más conocido ocurre en las arterias, que provoca un aumento de la presión sanguínea. Lo que propone es desarrollar fármacos capaces de llevar hacia atrás esas reacciones.
- Daño en las mitocondrias. El ADN mitocondrial no tiene los potentes mecanismos de reparación del ADN nuclear, por lo que los daños aquí se producen antes y llevan al individuo a problemas metabólicos y energéticos. Para evitarlo propone algo también muy radical e invasivo: llevar los genes de las mitocondrias al núcleo, salvando todos los obstáculos que ello va a implicar.

También es un buen indicador que la tasa de aumento de la esperanza de vida es actualmente de 3 meses por cada año que pasa, y sigue creciendo. Cuando llegue a 12 meses por cada año que pasa, las personas serán inmortales. Y muchos investigadores dicen que eso ya ha ocurrido de una u otra forma, por ejemplo, que la primera persona en alcanzar los 1000 años de edad ya ha nacido. Claro que hay que entender que son solo estimaciones estadísticas.

CAMPEÓN EN REPRODUCCIÓN

La pregunta tiene trampa. No estoy pensando en ningún ser vivo convencional, sino en los computadores. En 1930 no existía ninguno. Hoy en 2017 se estima que hay más computadores que personas, lo que se acentúa si incluimos los dispositivos de cómputo embebidos como teléfonos, electrodomésticos, automóviles, etc. Claro que los computadores no se reproducen solos sino que necesitan convencer a los humanos para que les ayuden. Y parece que nos han convencido muy bien.

¿Cuánto falta para que los computadores se reproduzcan sin nuestra ayuda? Muy poco porque las fábricas están ya automatizadas. Y algo importante va a cambiar cuando ello ocurra, ¿verdad?

COMUNICACIONES EN 2D

El problema no tiene solución en dos dimensiones y lo más que se puede hacer es poner dos cables (Figura 112). El tercero es imposible. Por otro lado, en tres dimensiones es trivial de resolver. Es más, cualquier problema de este tipo tiene solución en tres dimensiones.

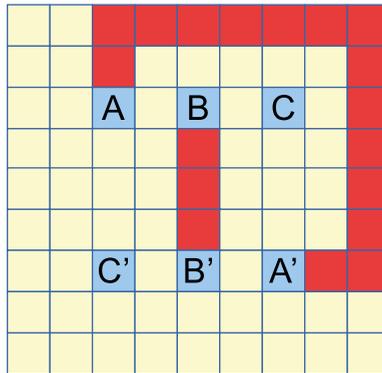


Figura 112. No hay solución

Cuando se diseñan circuitos impresos electrónicos (las tarjetas de color verde que hay dentro de nuestros computadores), este problema es muy habitual. Hay miles de puntos de conexión (en vez de los 6 de este problema) y hay que enrutar pistas de cobre para conectarlos. Como el circuito es plano, ello es imposible. Lo que se hace para solucionarlo es utilizar circuitos impresos de múltiples capas, típicamente cuatro: dos para llevar a todos los sitios la alimentación de 0 v y +3.3 v. Y las otras dos para las conexiones, de modo que cuando se agota el espacio en una, se puede hacer un agujero metalizado que nos conecte con la otra. Y a veces ni siquiera ello es suficiente, en circuitos con muy alta densidad de conexiones, por lo que se fabrican con más capas. Desde luego, en 3D no hay ninguna limitación.

EN LA CONSULTA MÉDICA

La respuesta a esto es muy personal. No sé lo que hará usted, pero yo lo tengo muy claro: prefiero al hermano de AlphaGo, no solo en medicina sino en cualquier profesión, incluyendo las mías que son profesor e ingeniero.

Si ya has estudiado alguna carrera relacionada con los computadores o la electrónica digital seguro que te has dado cuenta que 15 es $16-1$; y 18 es $16+2$. Dieciséis es un bonito número en base dos (1000_2). Entonces $15 \times 18 = (16-1) \times (16+2) = 16 \times 16 + 16 - 2$ y estas operaciones son triviales en binario, gracias a que multiplicar por dieciséis equivale a añadir tres ceros al otro número. El resultado es $256 + 16 - 2 = 255 + 15 = 270$.

Se puede multiplicar comenzando por la izquierda (FreeTestPrep, 2012). Y hay otras formas muy rápidas, que las emplean las personas que quieren deslumbrar con sus cálculos de cabeza (Kumar, 2012).

Cuantas más formas tengas de resolver un problema, cuantas más relaciones fabriques con otros conceptos, incluso de áreas muy dispares, más inteligente eres.

SALVAR EL FOSO

La solución es aprovechar una esquina, como puede verse en la figura 115.

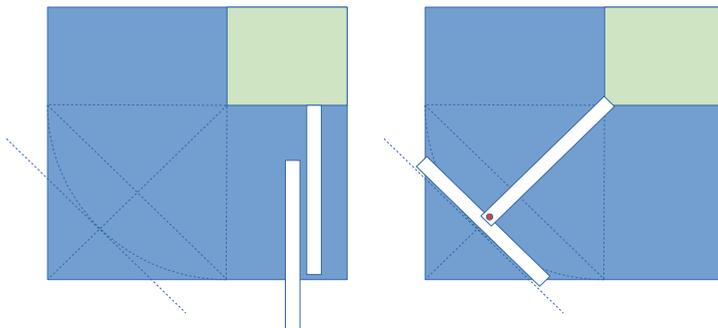


Figura 115. Izquierda: las dos tablas no parecen alcanzar. Derecha: hay un punto cerca de cada esquina (marcado en rojo) que es alcanzable por las dos tablas dispuestas en diagonales cruzadas, en una configuración estable que permite pasar al otro lado del foso

NIM

Lo primero que debes hacer es contar el número de monedas de cada fila y expresarlo en binario. Y luego contar la cantidad de '1' en cada columna de los números binarios. En la figura 116 se pueden ver tres ejemplos.

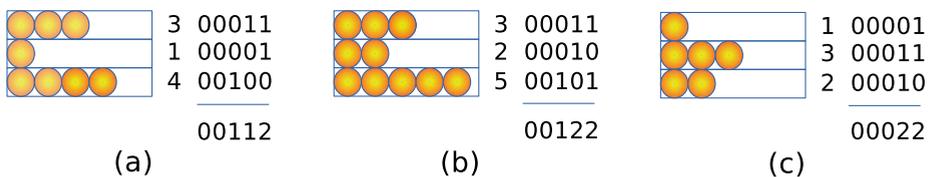


Figura 116. Tres ejemplos de NIM

Se dice que un juego de NIM está equilibrado si es par la cantidad de ‘1’ en todas las columnas de los números binarios. Y desequilibrado en caso contrario. En (a) y (b) el juego está desequilibrado, mientras que en (c) está equilibrado.

Es fácil de ver que, dado un juego equilibrado, cualquier jugada posible lo transforma en desequilibrado.

Por ello, la estrategia ganadora consiste en recoger monedas de modo que, después de ello, el juego quede equilibrado. Si inicialmente el juego está desequilibrado, el primero en jugar es el que gana, siempre que use esta estrategia. Si el juego está equilibrado, es el segundo jugador el que gana.

Mantener un invariante, el equilibrio en el juego, conduce a la victoria.

DETECTAR ENGAÑOS

Este experimento se le atribuye al psicólogo Peter Wason (en Pinker, 1997, p. 356) y a John Tooby y Leda Cosmides (en Boyd y Silk, 2004, p. 530). La mayoría de la gente falla en la parte a) pero acierta sin problemas en la b), lo cual indica que nuestro cerebro está adaptado a resolver problemas sociales, pero no abstractos, a pesar de que se trata del mismo problema, una implicación de tipo $P \Rightarrow Q$.

La respuesta para a) es que hay que girar la carta D y la 7.

La respuesta para b) es que hay que preguntar la edad al bebedor de cerveza; y qué está bebiendo a la persona de 16 años.

LAS TRES PUERTAS

La mayoría de la gente piensa que da igual si permanecemos con la puerta inicialmente escogida o nos cambiamos a la puerta que no abrió el presentador, pues la probabilidad de que el premio esté en una puerta cualquiera es $\frac{1}{3}$. Sin embargo, no es así: es mejor cambiar de puerta y enseguida demostraremos por qué.

Cuando leí este problema por primera vez en Hoffman (2000), yo también me equivoqué y no entendí la explicación para la solución que daban. Y mi ofuscación era tal que escribí un programa en Ruby simulando el concurso, ejecutándolo muchas veces, con el premio colocado al azar en cualquier puerta, lo mismo que las elecciones del concursante, y contando al final en cuantas ocasiones se llevaba el premio. El *software* indicaba que era mucho mejor cambiar de puerta. Entonces pensé que mi *software* tendría algún *bug*. En fin, me demoré varias horas en entender mi error.

El problema lo planteó por primera vez en 1990 Marilyn vos Savant, una persona con un coeficiente intelectual de los más altos en el mundo, a pesar de que no tiene estudios universitarios. Y generó mucha polémica, pues la demostración tampoco era fácil de entender. Espero que esta sí lo sea.

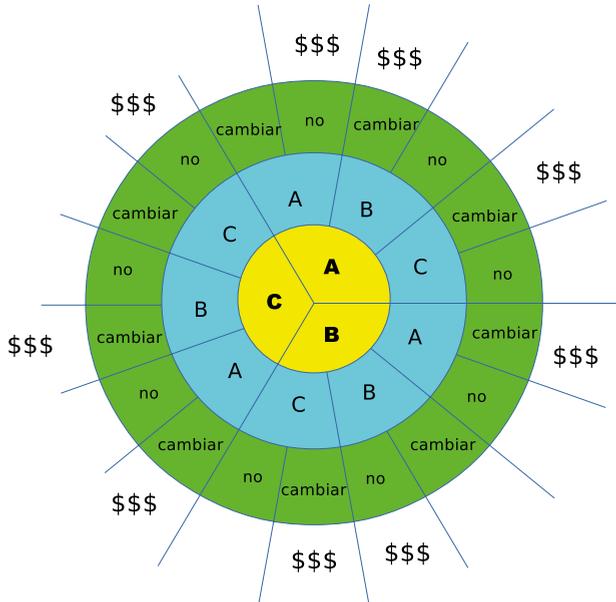


Figura 117. Solución al problema de las 3 puertas A,B,C

En el círculo central de la figura 117, en color amarillo están las posibles ubicaciones del premio (puertas A, B o C). En el siguiente círculo, en color azul, están las posibles elecciones iniciales de puerta (A, B o C). En el círculo externo, de color verde, están las opciones de cambiar o no cambiar de puerta. Por último, por fuera de los círculos se indican con \$\$\$ los casos donde el concursante obtiene el premio.

Conclusión:

- En 6 casos, cambiar de puerta me conduce a premio (\$\$\$).
- En 3 casos, no cambiar de puerta me conduce a premio (\$\$\$).

Por tanto, es mejor cambiar de puerta, ya que la probabilidad de obtener premio es el doble que si me quedo con la puerta inicial.

Hay algo muy sutil que nos cuesta entender: aunque abrir una puerta vacía puede parecer que no cambia las cosas pues no aporta ninguna información, resulta que sí lo hace. Porque el presentador de televisión no elige al azar, sino que abre una puerta que sabe perfectamente que está vacía, a la vez que no puede abrir la que hemos elegido nosotros. Nos está dando información.

También puede entenderse usando el teorema de Bayes. Supongamos, sin pérdida de generalidad, que inicialmente elijo la primera puerta.

$$p(\text{premio en primera puerta}) = \frac{1}{3} \quad \text{Ec. 40}$$

El presentador abre la segunda puerta. Entonces el premio quizás esté en la primera o en la tercera. La probabilidad de la ecuación 40 no cambia, pero ahora sabemos que la segunda puerta no tiene premio. Por tanto:

$$\begin{aligned} & p(\text{premio en tercera puerta} | \text{segunda puerta sin premio}) \\ = & \frac{p(\text{segunda puerta sin premio} | \text{premio en tercera puerta}) \cdot p(\text{premio en tercera puerta})}{p(\text{segunda puerta sin premio})} \quad \text{Ec. 41} \\ = & \frac{1 \cdot \frac{1}{3}}{\frac{2}{3}} = \frac{1}{2} \end{aligned}$$

Resumiendo, es mejor cambiar a la tercera puerta porque la probabilidad de premio es mayor:

$$\begin{aligned} p(\text{premio en primera puerta}) &= \frac{1}{3} \\ p(\text{premio en tercera puerta} | \text{segunda puerta sin premio}) &= \frac{1}{2} \quad \text{Ec. 42} \end{aligned}$$

APLANANDO LA CIRCUNFERENCIA

Antes de realizar ningún cálculo hay que demostrar que la serie de N semicircunferencias converge hacia la línea conforme N tiende a infinito. Y eso no ocurre, pues la pendiente de las semicircunferencias en los puntos que interseca con la línea no varía al aumentar N . Por tanto no hay convergencia, de modo que el resto de cálculos no tienen ningún sentido.

PUNTO FIJO DE PALMIRA A TULUÁ

Aunque parezca imposible, sí se puede demostrar que existe al menos un punto en la carretera por el que pasó Luisa a la misma hora del lunes y del martes. El punto en sí no se puede determinar, pues no hay datos suficientes, pero su existencia es fácil de ver si realizamos un gráfico del espacio recorrido respecto a la hora, y superponemos los trayectos de los dos días. En la figura 118-a podemos ver un viaje típico, en color verde de Palmira a Tuluá, y en rojo de Tuluá a Palmira. Dado que la función es continua, no queda alternativa: debe haber un punto de intersección entre los dos recorridos. En la figura 118-b observamos que Luisa se detuvo mucho tiempo apenas salió

de Tuluá, quizás por un pinchazo, y luego tuvo que coger más velocidad para llegar a las 13:00 a Palmira, pero también hay un punto de intersección entre los dos recorridos. En la figura 118-c vemos que tuvo que retroceder al poco de salir de Tuluá, quizás se le olvidó saludar a alguien, pero eso es irrelevante pues al final también existe un punto de intersección entre los dos recorridos. Y en la figura 118-d vemos que a causa de esos retrocesos no solo hay un punto de intersección sino tres.

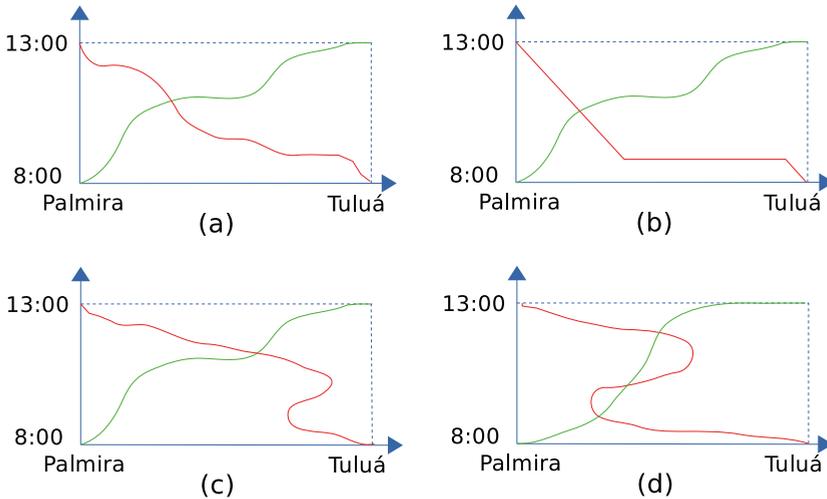


Figura 118. Algunas formas de hacer el recorrido de ida y vuelta

Es fácil de ver que independientemente de la velocidad que lleve, siempre existe al menos un punto de la carretera donde estuvo a la misma hora el lunes y el martes. A ese punto se le llama “punto fijo”.

ARRUGANDO PAPELES

Aprovecha que la caja tiene muchos más papeles y dibuja en color rojo la silueta del primer papel arrugado sobre el segundo papel de la caja, que está liso. Es así como proyectar la sombra, verticalmente.

Sobre ese segundo papel liso (en 2D) habrá ahora una línea roja, seguramente muy quebrada, pero es continua y es cerrada.

Toma después ese segundo papel y vuélvelo a arrugar exactamente de la misma forma como hiciste con el primero. Sí, ya sé que eso es muy difícil en la práctica, pero continuemos con el experimento mental, pues en principio puede hacerse.

Ahora que está arrugado, el papel contiene una curva roja en 3D, pero sigue siendo cerrada y continua. No hay “islas” ni discontinuidades porque

el papel solo se ha doblado, sin llegar a romperlo en ningún sitio. Arroja este segundo papel sobre la caja, exactamente en el sitio y orientación que lo hiciste con el primer papel.

Proyecta verticalmente esa línea roja sobre el tercer papel liso, dibujando sobre él una nueva línea roja. Obviamente, el área que encierra esta nueva línea estará dentro de la primera. El contorno rojo se ha contraído.

Repite este proceso un gran número de veces. El contorno rojo se irá contrayendo cada vez más hasta reducirse a un punto. Ese es el punto que no ha cambiado de sitio. Ese es el punto fijo.

Puede haber más de un punto fijo. De hecho, o hay uno o hay infinitos¹⁰⁴. Por ejemplo, podría ser un área irregular que te has empeñado en mantener intacta mientras arrugabas el resto del papel. O también un segmento de línea. Hay un caso especial a analizar, donde la línea cerrada roja no converge a un punto, y es si la zona que te has empeñado en mantener lista y en su mismo sitio, también la has rotado horizontalmente intentando que los puntos no queden donde estaban originalmente (si se trata de un círculo, un cuadrado u otra figura geométrica que tenga simetría de rotación), o la has dado la vuelta verticalmente (para lo que se requiere que tenga simetría especular). En ambos casos habrá un centro de rotación o un eje de giro, que son los que no han cambiado de sitio. Queda claro que en todos los casos hay al menos un punto fijo.

LA VELA

Este problema es muy viejo y de autor desconocido, aunque se lo escuché por primera vez a mi hermano Santi.

La solución es poner la vela horizontal y encenderla por ambos extremos a la vez. Cuando toda se haya consumido, habrá transcurrido media hora.

La razón de ello es la siguiente: al arder la vela por ambos extremos habrá un punto que alcancen las dos llamas simultáneamente. Ese es el punto fijo. En ese momento, toda la vela se habrá consumido. Ese punto no necesariamente está justo en la mitad, puesto que la vela no es uniforme y cada llama avanzará a diferentes velocidades. Sin embargo ese punto se alcanza exactamente en media hora. Entendamos por qué: supongamos que debido a las irregularidades de la vela, una llama alcanza ese punto en 43 minutos. Como

¹⁰⁴ Si hubiera dos puntos, entonces el segmento recto que los une también estaría conformado en su totalidad por puntos fijos. Porque en caso contrario habría que arrugar el segmento, y ello implicaría un acortamiento de la distancia entre los dos puntos iniciales, y para ello al menos uno se habría tenido que mover del sitio, por lo que no podrían ser puntos fijos ambos.

la otra llama llega a la vez, también en 43 minutos, entonces si hubiera una única llama recorriendo toda la vela, se demoraría $43+43=86$ minutos en consumirla, lo cual contradice el enunciado, que dice que la vela se consume en exactamente una hora. La única forma de no contradecir el enunciado es que ambas llamas alcancen el punto fijo en 30 minutos, es decir, media hora.

PARA SABER MÁS

La mayoría de los autores de estos problemas son anónimos. Me los contaron frente a unas cañas o proceden de revistas de entretenimiento donde no aparece ninguna referencia.

Los problemas de ingenio son un pasatiempo divertido, despiertan nuestra creatividad, sirven para aumentar nuestro coeficiente intelectual y mantienen joven nuestro cerebro. Los seleccionados aquí, además, iluminan algunos de los temas que se abordan en este libro.

A continuación indico cuatro libros donde hay recopilaciones de estos problemas.

- Juan José Rivera Gómez (1981). *Comecocos I*. Madrid: Editorial Álamo.

Un libro muy viejo con muchos problemas de ingenio, la mayoría sencillos, pero contiene también algunas joyas como “La mayor toca el piano”, que ya mencioné en el libro anterior. Si lo resolviste no será muy difícil entender este otro, que apareció en la revista *Investigación y Ciencia* (la versión en español de *Scientific American*), sección Juegos Metamágicos, hace un montón de años (febrero de 1980), probablemente de Martin Gardner:

Hay dos números enteros M y N , que están entre 1 y 20, no necesariamente distintos, ambos excluidos. Y hay dos matemáticos S y P . Al matemático S se le dice cuánto vale la suma $M+N$, mientras que al matemático P se le dice cuánto vale el producto $M*N$. Después se desarrolla la siguiente conversación:

Matemático S : No sé cómo vas a averiguar mi suma.

Matemático P : Ya sé tu suma.

Matemático S : Entonces yo ya sé tu producto.

¿De qué números M y N se trata?

- Lewis Carroll (1979). *Matemática demente*. Barcelona: Tusquets Editores.

Del mismo autor de “Alicia a través del espejo”. Hay un cuento espectacular, “Lo que le dijo la tortuga a Aquiles”, que demuestra que incluso para aplicar la lógica se necesita sentido común. Si hay reglas, se requieren me-

tarreglas, para que las primeras no sean abusadas. Y hay luego unos pocos problemas de ingenio que hacen reflexionar sobre cosas cotidianas como: ¿por qué los espejos invierten izquierda y derecha, pero no arriba y abajo?

- Martin Gardner (1992). *Inspiración Ajá*. Barcelona: Editorial Labor.

Este libro es técnicamente el mejor, el más complejo y que toca muchos aspectos de las matemáticas, no solo la lógica.

- Raymond Smullyan (1978). *¿Cómo se llama este libro?. El enigma de Drácula y otros pasatiempos lógicos*. Madrid: Prentice Hall.

Este libro también es muy bueno, pero se centra únicamente en la lógica.

REFERENCIAS

Libros, artículos y enlaces web

Arthur, W. B. (1994). Inductive Reasoning and Bounded Rationality (The El Farol Problem). In *American Economic Review - Papers and Proceedings*, 84(406), pp. 1-11.

Boyd, R. y Silk, J. B. (2004). *Cómo evolucionaron los humanos*. Barcelona: Editorial Ariel.

de Grey, A. D. N. J. (2017). Revertir el envejecimiento mediante la reparación de daños moleculares y celulares. En *El próximo paso: la vida exponencial*. España: BBVA, OpenMind.

Hoffman, P. (2000). *El hombre que solo amaba los números. La historia de Paul Erdős y la búsqueda de la verdad matemática*. Barcelona: Ediciones Granica.

Hofstadter, D. R. (1979). *Gödel, Escher, Bach: un eterno y grácil bucle*. Barcelona: Tusquets Editores.

Pinker, S. (1997). *Como a mente funciona*. São Paulo: Editora Schwarcz.

RP_Wiki (2017). Richard's Paradox – Variation: Richardian Numbers. Recuperado el 21 de agosto de 2017. Disponible en: <https://goo.gl/7F1r6X>

von Savant, M. (1990). *Game Show Problem*. Recuperado el 12 de abril de 2017. Disponible en: <https://goo.gl/VwqdU5>

Películas y videos

FreeTestPrep (2012). *Speed Multiplication (1 digit) from Left to Right*. Recuperado el 2 de octubre de 2016. Disponible en: <https://goo.gl/FjJa4f>

Kumar, P. (2012). *Fast Multiplication Trick 5 - Trick to Directly Multiply the Big Numbers*. Recuperado el 2 de octubre de 2016. Disponible en: <https://goo.gl/Rme7xS>

GLOSARIO

AC: Autómata celular. Un grafo regular e infinito, donde cada nodo está conectado bidireccionalmente solo con sus vecinos.

FSM: *Finite State Machine*, máquina de estados finitos. Grafo dirigido con una única marca indicando el estado activo. La marca puede moverse a otro estado a través de algún arco saliente de ese estado, si se cumple la condición indicada en el arco. Si hay más de una marca activa, se llama Red de Petri.

Estocástico: una secuencia es estocástica si cada término es imposible de predecir conociendo los anteriores. Las secuencias estocásticas también se pueden llamar al azar o no-deterministas. No es lo mismo que aleatorio, aunque mucha gente los confunde.

Aleatorio: una secuencia es aleatoria si no se puede comprimir. No es lo mismo que estocástico, aunque mucha gente los confunde.

Seudoaleatorio: es similar a un proceso caótico digital, con horizonte de predicción de una unidad de tiempo (o sea, solo se puede predecir la salida actual, pero no las siguientes, a partir de todas las entradas pasadas y de la fórmula) y donde permanece oculta la fórmula de conversión de entradas en salidas. El hecho de que sea digital implica que no hay ruido en las entradas, de modo que estrictamente hablando no es caótico. Pero la idea de un cortísimo horizonte de predictibilidad permanece.

Caótico: la definición rigurosa puede verse en el correspondiente capítulo, pero aproximadamente se puede decir que un proceso es caótico si es muy sensible a sus entradas, es decir, con un minúsculo cambio de la entrada, la salida cambia mucho. Eso hace que sean difíciles de predecir y suele haber un horizonte de predicción a partir del cual los errores acumulados hacen imposible anticipar el futuro del sistema.

Determinista: una secuencia de datos (o un proceso) es determinista si las sucesivas salidas (o estados) están completamente determinadas por las salidas anteriores (o estados anteriores). Los procesos deterministas convierten sus entradas en salidas por medio de una fórmula o un algoritmo que no contiene ningún elemento de azar. De modo que si se conocen las condiciones iniciales, se pueden predecir las salidas sucesivas. Y si las entradas se repiten, también lo hacen las salidas.

Fractal: un objeto geométrico cuya dimensión de Hausdorff-Besicovitch no coincide con su dimensión topológica. Informalmente se dice cuando un objeto tiene infinita rugosidad y autosemejanza en todas las escalas.



Universidad
del Valle

Programa ditorial

Ciudad Universitaria, Meléndez

Cali, Colombia

Tel.: (572) 321 2227

321 2100 ext. 7687

<http://programaeditorial.univalle.edu.co/>

E-mail: programa.editorial@correounivalle.edu.co