



Convergencia de experimento y teoría en los procesos de invención e innovación

(Convergence between experiment and theory in the processes of invention and innovation)

David CASACUBERTA*, Anna ESTANY

Universitat Autònoma de Barcelona, Spain

RESUMEN: Este artículo parte del debate en filosofía de la ciencia entre la tradición teórica y la experimental, y muestra su relación con el estudio de los procesos de innovación e invención en ciencia, cruzando así los planteamientos de análisis más teóricos de la filosofía de la ciencia con cuestiones más relacionadas con la filosofía de la tecnología y la ciencia aplicada. De esta manera, analizamos la interrelación entre experimento y teoría en los procesos de invención e innovación y conectamos los campos de la ciencia teórica y la aplicada, mostrando la continuidad entre ambas. Así, podemos mostrar también cómo en ciencia hay siempre dependencia mutua de teoría y experimentación, y cómo esa dependencia es extrapolable también a los procesos de innovación e invención.

Tomando como punto de partida el debate en torno a las tradiciones teóricas y experimentales, veremos hasta qué punto los argumentos que cuestionan las tradiciones teóricas y apuestan por las tradiciones experimentales encajan con los fenómenos de invención e innovación. El caso que vamos a tomar como referencia para aplicar este análisis es el de «aprendizaje automático», como una rama de los algoritmos computacionales diseñados para emular la inteligencia humana aprendiendo del entorno. Este campo es relevante pues, a pesar de su naturaleza eminentemente teórica —en substancia es matemática aplicada—, presenta toda una serie de características que lo hacen muy afin al análisis desde las tradiciones experimentales.

PALABRAS CLAVE: Experimentación, innovación, invención, aprendizaje automático.

ABSTRACT: *This article starts from the debate in philosophy of science between the theoretical and the experimental traditions, and it aims to show its relation with the study of innovation and invention processes in science, thus crossing the most theoretical approaches of the philosophy of science with issues more related to the philosophy of technology and applied science. In this way we analyze the interrelation between experiment and theory in the processes of invention and innovation and connect the fields of theoretical and applied science, thus showing the continuity between them. That way, we can also show how in science there is always mutual dependence on theory and experimentation, and how that dependence can also be extrapolated to the processes of innovation and invention.*

Taking as starting point the debate around the theoretical and experimental traditions, we will see to what extent the arguments that question the theoretical traditions and opt for the experimental ones fit with the phenomena of invention and innovation. The case that we are going to take as a reference to apply this analysis is that of «machine learning», as a branch of computational algorithms designed to emulate human intelligence by learning from the environment. This field is relevant because, in spite of its eminently theoretical nature—in substance it is applied mathematics—, it presents a whole series of characteristics that makes it very similar to the analysis from the experimental traditions.

KEYWORDS: *Experimentation, innovation, invention, machine learning.*

* **Correspondencia a/Correspondence to:** David Casacuberta. Departamento de Filosofía, Edificio B, Campus de la UAB (08193 Bellaterra-Cerdanyola del Vallès), España – david.casacuberta@uab.cat – <http://orcid.org/0000-0001-7119-9342>

Cómo citar/How to cite: Casacuberta, David; Estany, Anna. (2019). «Convergencia de experimento y teoría en los procesos de invención e innovación»; *Theoria. An International Journal for Theory, History and Foundations of Science*, 34(3), 373-387. (<https://doi.org/10.1387/theoria.17921>).

Recibido/Received: 09 junio, 2017; Versión final/Final version: 16 abril, 2018.

ISSN 0495-4548 - eISSN 2171-679X / © 2019 UPV/EHU



Esta obra está bajo una licencia
Creative Commons Attribution 4.0 International License

La relación entre teoría y experimento ha estado siempre presente en la filosofía de la ciencia, especialmente en el debate entre tradiciones teóricas y experimentales. La cuestión está en el papel de la teoría y el experimento en la investigación científica y en el predominio de una tradición sobre otra. Históricamente, tanto para filósofos como historiadores de la ciencia ha predominado la idea de que la experimentación es subsidiaria de la teoría, pero hace ya unas décadas que esta concepción se ha puesto en cuestión. Por lo que respecta a las categorías de invención y de innovación históricamente se las ha asociado a la tecnología y a las ciencias aplicadas. Sin embargo, actualmente dichas categorías engloban la idea de cambio en cualquiera de los ámbitos, desde los científicos a los culturales y desde la tecnología a las formas de vida cotidiana, siendo también un referente para cambios metodológicos y epistemológicos.

El objetivo de este artículo es analizar la interrelación entre experimento y teoría en los procesos de invención e innovación. Queremos mostrar cómo en ciencia hay siempre dependencia mutua de teoría y experimentación, y cómo esa dependencia es extrapolable también a los procesos de innovación e invención. Tomando como punto de partida el debate en torno a las tradiciones teóricas y experimentales, veremos hasta qué punto los argumentos que cuestionan las tradiciones teóricas y apuestan por las tradiciones experimentales encajan con los fenómenos de invención e innovación. El caso que vamos a tomar como referencia para aplicar este análisis es el de «aprendizaje automático»,¹ como una rama de los algoritmos computacionales diseñados para emular la inteligencia humana aprendiendo del entorno. En resumen, el objetivo es partir de un debate en filosofía de la ciencia (tradiciones teóricas *versus* tradiciones experimentales) y contrastarlo con aportaciones al campo de la invención e innovación que, preminentemente, proceden de la filosofía de la tecnología y la ciencia aplicada. Precisamente la novedad del artículo consiste en conectar dos campos aplicando modelos de la filosofía de la ciencia al ámbito de la ciencia aplicada.

En primer lugar, abordamos los principales ejes por los que discurre el debate sobre tradiciones teóricas y experimentales; en segundo lugar, examinamos las ideas básicas sobre innovación e invención, tomando como marco de referencia las tesis de las tradiciones teóricas y experimentales; en tercer lugar, analizamos el caso de aprendizaje automático desde el análisis y la discusión en las secciones anteriores y, finalmente, hacemos una serie de consideraciones a partir de los análisis realizados.

1. *Tradiciones teóricas versus tradiciones experimentales*

El debate sobre el papel de la teoría *versus* el papel del experimento es antiguo en filosofía de la ciencia. No cabe duda que en la actividad científica se encuentran tanto científicos teóricos como experimentales (Artigas 1989) lo cual es, en cierta forma, un reconocimiento de dos actividades diferentes aunque mutuamente necesarias. Sin embargo, la filosofía de la ciencia se ha centrado en los modelos teóricos, dejando a la experimentación

¹ Traducimos así el término «machine learning», aunque actualmente no hay un consenso claro sobre como traducir esta palabra. La tendencia más establecida es todavía dejar el término en inglés, pero consideramos que «aprendizaje automático» captura el concepto suficientemente bien y es el término en castellano más usado.

un papel subsidiario. Para la mayoría de las escuelas, tanto de la tradición heredada como de la visión kuhniiana de la ciencia, la experimentación ha estado en función de la teoría, ya sea inspirada por ella o al servicio de la misma, pero en cualquier caso, sin vida propia (Hacking 1996). Para esta concepción de ciencia, el experimento es considerado como verificador de teorías o como una forma de «experimentación exploratoria», tal como lo consideró el empirismo lógico de Carl Hempel, o falsador de las mismas desde un punto de vista popperiano. Para este análisis hemos incorporado la propuesta de algunos de los autores representativos que han cuestionado el papel prioritario de la teoría como ejemplo de esta propuesta, pero queremos dejar claro que no hemos pretendido hacer un análisis de este debate ni dar un panorama exhaustivo del mismo.

Hay una cuestión que debemos aclarar desde el principio y es si al hablar de experimentación, y quizás también de teoría, nos referimos a un enfoque teórico o aplicado. Digamos que, tradicionalmente, la filosofía de la ciencia se ha centrado en la ciencia pura o básica, y el debate alrededor de las tradiciones teóricas *versus* las experimentales se refiere a la ciencia básica. Sin embargo, no cabe duda que también en las ciencias aplicadas hay teoría y experimentación, posiblemente con metodologías distintas, y las cuestiones que aquí se analizan competen también a las ciencias aplicadas. Profundizar en las posibles diferencias va más allá de los objetivos de este trabajo.

Ian Hacking (1996) da un giro en la filosofía de la ciencia reivindicando el papel del experimento en la práctica científica y señala que: «la filosofía debía empezar a reflexionar sobre lo que comenzó por allá en el siglo XVII. La aventura que entonces se inició, y que puso por base la experimentación, fue llamada filosofía experimental» (Hacking 1996, 52). Su propuesta no es tanto eliminar las teorías como representación del mundo sino abandonar el esquema jerárquico en el que el experimento está al servicio de la teoría. Reconoce que podemos encontrarnos casos, tanto en que el experimento prima sobre la teoría (tal sería el caso de Priestley en química o Faraday en física), como casos en que sería difícil hacer avanzar la ciencia sin tener en cuenta determinados modelos teóricos (por ejemplo, en la época de Liebig no podían resolverse problemas de química orgánica sin partir de determinados modelos teóricos de la química). De todo ello podemos concluir que el peso de la teoría o la experimentación en la investigación científica no parece que sea una cuestión que podamos decidir *a priori* sino que, como muestran los ejemplos de Hacking, dependerá del momento histórico en el que se encuentre la disciplina en cuestión.

Por su parte, Mercedes Iglesias (2004) hace referencia a la importancia de las prácticas experimentales para mostrar el giro necesario en filosofía de la ciencia y el cambio de la tradicional relación entre teoría y experimento. El giro hacia la práctica en filosofía de la ciencia obliga a que los temas de racionalidad, objetividad, verdad y mundo dejen de ser tratados desde la teoría y a que se redefinan nuevos problemas filosóficos, promoviendo una nueva imagen de la ciencia. Por otra parte Anna Estany (2007) avanza en la búsqueda de resultados empíricos de las ciencias cognitivas que permitan afianzar el sentido de las prácticas experimentales.

Con las investigaciones de estudios de caso, los filósofos de la ciencia quieren poner de manifiesto la existencia de la carga experimental de la teoría. Tal es el caso de Peter Galison (1987), Andrew Pickering (1995) y David Gooding, Trevor J. Pinch y Simon Schaffer (1989) y Sergio Martínez (2003) entre otros, que avanzan hacia una nueva imagen de ciencia. Javier Ordoñez y José Ferreiros (2002) consideran que la miseria del teoreticismo está en reducir la riqueza y la complejidad del proceder científico a un asunto de mera elabora-

ción conceptual dejando de lado la riqueza de conocimiento que se esconde detrás de las prácticas experimentales. Reconocer su importancia y la validez de las mismas en la constitución de la ciencia, su función independiente de la teoría o en equilibrio con ella y su papel más allá del verificacionista o falsacionista que usualmente se le ha otorgado, constituye el fundamento de este campo de investigación de la filosofía de la ciencia.

Otro aspecto que resulta importante desde la filosofía de las prácticas experimentales es el tipo de discurso que hay en la propia experimentación, que no se corresponde con el que en la concepción heredada se le asigna a la deducción (Galison 1987). Es necesario admitir que en la acción hay pensamiento, lo cual implica romper y superar el dualismo cartesiano que divide mente y cuerpo, naturaleza y cultura. Hay un lenguaje diferente que se expresa en la actividad experimental y del cual surgen pensamientos e ideas que posteriormente se articulan conceptualmente. El conocimiento experimental está presente al diseñar y construir aparatos, pero también en la manipulación de entidades y en la creación de fenómenos. Hemos de reconocer entonces que en la actividad experimental hay una riqueza conceptual que no había sido reconocida ni valorada en su justa dimensión.

En este sentido Iglesias señala que «la naturaleza no se nos muestra ella por sí sola» (Iglesias 2004, 11): ella se abre, se despliega, según lo imponga la manera a la que fue sometida en una acción específica. Finalmente debemos reconocer que la nueva imagen de ciencia está orientada por una relación entre la teoría y el experimento mucho más profunda y con mayor riqueza conceptual de la que la mayoría de los filósofos e historiadores han mantenido hasta el momento, pero las nuevas corrientes filosóficas promueven la importancia de la «vida propia» que posee la experimentación en la construcción de la ciencia, y que como afirma Iglesias:

Cuando se asume la perspectiva de estudios de casos, de puntos específicos de la historia de la ciencia, se encuentra que no es la teoría la que siempre ha guiado la ciencia, que la relación entre teoría y experimento ha sido diversa y no unitaria como han querido mostrarnos la historia y la filosofía de la ciencia. (Iglesias 2004, 10)

Hacer de la experimentación una plataforma de conocimiento contribuye a un cambio en la imagen de la ciencia. La manera de presentar los experimentos no debe ser solamente descriptiva o narrativa para reforzar el papel de las teorías, sino que se debe avanzar hacia la caracterización de experimentos que involucren problemáticas y que tengan una riqueza conceptual en sí mismos (vida propia), crear necesidades donde el experimento «hable» y se comunique, crear situaciones específicas donde la naturaleza se «despliegue» y muestre comportamientos, esto es, hacer de la experimentación una actividad humana en el sentido de Hacking.

2. *Los procesos de invención e innovación*

La invención y la innovación, aunque no en estos términos, podemos rastrearla desde el principio de los tiempos y son anteriores a la ciencia como la entendemos actualmente. Las técnicas o artes de cultivar, curar, educar, contar, etc. fueron incorporando los conocimientos científicos a fin de alcanzar de la mejor manera posible sus objetivos. Actualmente estas artes se han convertido en lo que Herbert Simon e Ilka Niiniluoto llaman «ciencias de diseño», cuyo objetivo no es describir el mundo sino transformarlo. Situándonos en el momento actual, y a pesar de que tenemos una idea intuitiva de las ideas de invención e inno-

vación, cuando se trata de definir las aflora su polisemia. Es por ello que vamos a examinar algunos de sus sentidos más significativos así como sus diferencias.

La idea de invención está ligada a la novedad en un ámbito determinado a lo largo de la historia de la humanidad. Una buena muestra de ello la encontramos en el libro de Jeremy Stan (2012), que presenta las 100 invenciones más notables, desde la alfarería 12000 años antes de J. C. hasta el CD en 1982. Los ejemplos son de lo más variado y abarcan contextos totalmente distintos, reflejando las vicisitudes de las invenciones a lo largo de los siglos. El elemento común sería la novedad, por lo que podríamos decir que invención es un concepto integrador de lo que ha supuesto cambios para los humanos. A modo de ilustración podemos señalar los siguientes: la agricultura (8000 antes de J.C.), el vino (6000 antes de J.C.), la moneda (siglo VII antes de J.C.), el jabón (2500 antes de J.C.), el cristal (2500 antes de J.C.), la catapulta (399 antes de J.C.), la escritura (3400 antes de J.C.), la música (1600 antes de J.C.), la metalurgia (4000 antes de J.C.) y un largo etcétera. Con estos ejemplos vemos que están incluidos ámbitos muy distintos que van de lo que se consideraría cambios sociales a elementos de la vida cotidiana, pasando por artefactos tecnológicos que inciden en la ciencia, la cultura y el arte. Indudablemente el papel de la experimentación y de la teoría no es la misma en todos estos ejemplos pero todos son considerados invenciones y en muchas de ellas la teoría es mínima y, en todo caso, no como la entendemos ahora.

Al abordar el análisis conceptual de la innovación hay que tener en cuenta que algunas definiciones ponen el énfasis en la distinción entre invención e innovación, en virtud de la cual consideran la invención como un logro y la innovación como la actualización del mismo. Desde esta perspectiva, invención e innovación están relacionadas y constituyen dos fases de un proceso más general cuya consecuencia sería el avance en un campo determinado. Tal es el caso de las definiciones siguientes:

- *Invención* (creación de una idea nueva) e *Innovación* (primera utilización de una idea nueva), ambas estrechamente relacionadas con la palabra «técnica» (Edgerton 2013, 15).
- Invención es un logro e innovación es un actualización (Florida 1990, en Carayannis, González y Wetter 2003, 116).
- Es importante distinguir entre *invención* o generación de una idea original e *innovación* o proyecto de convertirlo en una producción útil (Roberts 1988, en Georgsdotir, Lubart y Getz 2003, 184).

Otros autores no hacen una distinción clara entre invención e innovación, lo cual no significa que no diferencien entre una idea original y su aplicación práctica. Tal es el caso, entre otros, de Arthur W. Arthur (2007) que considera la posibilidad de que una invención no logre convertirse en tecnología o resolver algún problema. Así, algunas invenciones pueden tener un recorrido largo y gradual, como en el caso de la energía solar, o una implantación rápida como pueden ser Internet y el móvil. Las razones pueden ser varias, desde la naturaleza de lo inventado a dificultades económicas. Es lo que llamaríamos «desfase entre la invención y la innovación» que, en términos de Arthur, sería la brecha entre un principio y la capacidad de cubrir una necesidad. Un ejemplo de desfase entre invención e innovación lo proporciona Stan (2012, 128) con la máquina de escribir, una invención que se intercala entre la escritura a mano y la informática. Se inventó a principios del siglo XVIII, pero no se desarrolló hasta finales del siglo XIX. En 1714 el inglés Henry Mill presentó una patente de

máquina de escribir y surgieron diversos prototipos durante la primera mitad del siglo XIX. En 1833 Xavier Progin tuvo la idea de reemplazar el cuadrante por teclas pero hubo que esperar a 1867 para que el impresor americano Christopher Scholes patentara su máquina de escribir (*typewriter*).

La aportación de Arthur (2007) es relevante en el análisis de la estructura de la invención. Considera que en la literatura sobre los cambios y novedades, sea en el ámbito que sea, la invención ha sido la menos estudiada. Su objetivo y la tesis que defiende pueden resumirse en las siguientes palabras:

Mi propósito es mostrar que la invención tiene una cierta lógica o estructura, y que, efectivamente, esta lógica explica por qué y cómo el proceso varía. Voy a argumentar que la invención es un proceso de vinculación de un propósito o necesidad con un efecto que puede ser explotado para satisfacerla. (Arthur 2007, 275)

A pesar de contemplar la posibilidad de que otros factores intervengan en un proceso de invención, en su modelo la tecnología es el factor desencadenante. En consecuencia, sus ideas sobre lo que entiende por tecnología son claves para comprender sus teorías sobre la invención. De entrada, define la tecnología como un medio para que los humanos logren un propósito y considera básico distinguir entre «concepto base» o «principio base» y «fenómeno».

En la argumentación que sigue, es importante que el lector tenga una clara distinción entre fenómeno y principio. Que la presión del aire cae con la altitud es un fenómeno físico; la idea de utilizar este efecto para medir la altitud constituye un principio. Un fenómeno es simplemente un efecto natural y, como tal, existe independientemente de los seres humanos y de la tecnología (...); no se le atribuye una «utilidad» determinada. En cambio, un principio (como voy a utilizar la palabra) es la idea de la utilización de un fenómeno con algún propósito, y existe en el mundo de los humanos y de cómo estos lo usan. (Arthur 2007, 277)

A partir de este marco teórico, Arthur define la invención como el cambio en alguno de los componentes tecnológicos que harán posible satisfacer la necesidad. Pero para que haya invención no es suficiente una mera modificación, sino que debe haber un cambio significativo, en el sentido de que logre el propósito utilizando un principio base diferente del utilizado hasta el momento (Arthur 2007, 278). Este es el criterio que marca la diferencia entre la mera modificación y la invención. El punto esencial de la invención es la conexión entre una necesidad y un principio que resuelva los problemas que han ido surgiendo.

En el modelo de Arthur el principio base sería lo equivalente a la teoría, ahora bien, la cuestión es si considera que dicho principio es prioritario y que la tecnología, y la utilización de la misma, son subsidiarios de dicho principio. No lo parece si tenemos en cuenta que para Arthur la invención es un proceso y no un acontecimiento. En este proceso contempla dos patrones: uno que se inicia por una necesidad y otro que se inicia a partir de un fenómeno. La necesidad puede tener un origen económico, social, militar, etc., aunque también puede responder a necesidades internas al propio proceso. Es decir, este proceso puede empezar por una necesidad y luego ver si hay un principio que puede satisfacerla.

Frente a una necesidad, surgirá algún individuo que intentará resolver el problema y al que llama «originador» en lugar de «inventor», término que considera excéntrico.² Trata a dicho originador como una mente individual, pero contempla la posibilidad de que intervengan varios originadores, por tanto, varias «mentes».³ En realidad, lo que dice puede entenderse que en el proceso de invención pueden intervenir varios agentes, ya que no parece que piense en una «mente colectiva».

En cualquier caso, los principios que darán lugar a la tecnología para cubrir una necesidad nunca surgen *ex nihilo*, sino de una serie de prácticas acumuladas y que, en algún momento, el originador hace la conexión entre un problema y un principio. En cuanto a los elementos que el originador utiliza para la invención tecnológica, Arthur introduce la idea de «funcionalidades» como aquellas acciones u operaciones de las que dispone el agente a partir de experiencias, conocimientos, metodologías, etc., y que pone en práctica para resolver un problema. El otro patrón de la invención es el que se inicia por la sugerencia de un fenómeno, aunque puede pasar tiempo hasta que consiga convertirse en tecnología y resolver un problema.

A partir de la relación entre teoría y experimento podemos hacer un paralelismo con el proceso de invención. Así como en las tradiciones teóricas primero es la teoría y luego el experimento, cuyo propósito es fundamentar la teoría en cuestión, lo equivalente en el caso de la invención, según el modelo de Arthur, primero sería el principio y luego la necesidad que intenta cubrir, sin embargo, Arthur contempla también la posibilidad de que el proceso de invención empiece por una necesidad y que se busque el principio que logre cubrirla. Es decir, que primero es la necesidad o el fenómeno y que la teoría está en función de satisfacer la necesidad o de buscar el principio que sustenta dicho fenómeno.

Si nos centramos en la innovación, lo común está en la utilidad práctica, la resolución de problemas, y de forma explícita o implícita, la conexión con la acción y la aceptación de la novedad por parte del colectivo o sociedad al que va dirigido. Estas características comunes son las que constituirían el concepto integrador de innovación. El énfasis en la utilidad está bien argumentado en Thomas Nickles (2003, 59), quien señala que la novedad tiene que ser útil, ya que la innovación es un término que expresa éxito y logro de un objetivo. Esta es una de las razones por las que Matti Sintonen (2009) plantea las dificultades de utilizar en las ciencias aplicadas los conceptos con los que caracterizamos los procesos creativos en las ciencias básicas o puras, ya que «la investigación aplicada es la búsqueda del conocimiento, donde el objetivo es, para emplear la caracterización autorizada de la OCDE desde hace unos 30 años, intentar aplicar los resultados de la investigación básica o incluso descubrir nuevos conocimientos que puedan tener una aplicación práctica inmediata» (Sintonen 2009, 215). Esta reflexión de Sintonen nos lleva a la conclusión que incluso si en las ciencias puras o descriptivas la experimentación tiene vida propia, en las ciencias de diseño con mucha mayor razón. Por este motivo, son las ciencias de diseño las que mejor enlazan

² El término inglés es *originator*, que hemos traducido por «originador». No parece relevante las diferencias entre los dos términos (*originator* e *inventor*). Si lo que quiere enfatizar Arthur es que el originador es el que tiene la idea, este es precisamente el significado de inventor. De todas formas, para la exposición del modelo de Arthur nos atenemos a su terminología.

³ Aunque no hace referencia a la «cognición distribuida», bien podría tomarse este modelo para abordar los casos en los que intervienen varias «mentes», que para Hutchins serían varios agentes.

con los procesos de innovación, ya que al tener como objetivo transformar el mundo, una invención que no se actualizara tendría poca trayectoria en cualquier proyecto práctico.

En el mismo sentido, Joseph S. Renzulli señala que «los objetivos de la ciencia nos dicen que el objetivo principal es añadir nuevos conocimientos a nuestra comprensión acerca de la condición humana, pero en un campo aplicado también hay un propósito práctico para la definición de los conceptos» (Renzulli 2003, 80). La primera parte de esta frase correspondería a la ciencia pura y es lo que podemos llamar la aportación de conocimiento sustantivo, la segunda pertenecería a la ciencia aplicada.

Una cuestión a tener en cuenta es dónde se pone el acento a la hora de valorar los factores más determinantes que generan procesos de innovación. Así pues, Dora Marinova y John Phillipmore (2003) señalan la importancia de la innovación tecnológica, que distinguen de la social, educativa u organizativa. En cambio Jon Sundbo (2003) toma el enfoque sociológico a pesar de que considera que el tema de la innovación es, en parte, económico. Por tanto, el autor apuesta por lo que denomina *strategic reflexivity*, un concepto que puede unificar las explicaciones económicas y sociológicas de la innovación, señalando que hay un elemento de verdad en ambos enfoques, lo que quiere decir que, si bien la estructura social influye mucho en el comportamiento de los individuos, las acciones individuales también determinan la estructura social (Sundbo 2003, 97).

Pensar que la introducción de la tecnología, además de factores sociales, educativos, organizativos, etc., ya es suficiente para dar prioridad a la experimentación por delante de la teoría, sería dar por supuesto de que la sociología, las ciencias de la educación y empresariales carecen de modelos teóricos. Por tanto, el debate sobre las tradiciones teóricas *versus* las experimentales afecta tanto a las ciencias naturales como sociales, y del mismo modo a las ciencias puras como aplicadas. Lo que sí es verdad es que en las ciencias aplicadas la parte experimental adquiere una relevancia especial porque están enfocadas desde el principio a resolver problemas.

3. *Análisis de caso: Aprendizaje Automático*

Al abordar la idea de aprendizaje automático nos enfrentamos a las posibles definiciones. Podríamos consumir todas las páginas de este número de la revista simplemente citando diferentes definiciones de aprendizaje automático. Uno de los manuales de referencia para el desarrollo de aplicaciones de aprendizaje automático más usados es el de Tom M. Mitchell (1997, 2), que define dicho concepto de la siguiente forma: «Afirmaremos que un programa de ordenador aprende de una experiencia E asociada a una clase de tareas T y una medida performativa P, si su capacidad performativa de las tareas T, medida a partir de P, mejora con la experiencia E.» Una definición más épica declara que el aprendizaje automático «es una rama en evolución de los algoritmos computacionales que están diseñados para emular la inteligencia humana aprendiendo del entorno que las rodea. Las técnicas basadas en aprendizaje automático han sido aplicadas de forma exitosa a diferentes disciplinas como reconocimiento de patrones, visión por ordenador, ingeniería espacial, finanzas, entretenimiento, así como biología computacional y aplicaciones médicas y biomédicas» (El Naqa y Murphy 2015, 4).

De todas formas, la más sencilla y elegante es la que aparece en Kevin B. Korb (2004) que afirma que «el aprendizaje automático estudia estrategias inductivas que pueden ser

ejecutadas por algoritmos». Korb ve una clara relación entre la forma de trabajo y objetivos del aprendizaje automático y la filosofía de la ciencia pues, según él, «[l]a filosofía de la ciencia estudia estrategias inductivas según aparecen en la práctica científica».

A manera de resumen, podríamos indicar que, finalmente, el aprendizaje automático busca simular algún tipo de proceso de aprendizaje, establecer cómo un algoritmo ha de recoger ciertos datos y analizarlos para conseguir establecer algún tipo de regularidad.

En cierta manera, simular usando modelos de aprendizaje automático, la forma en que un científico genera una hipótesis a partir de la evidencia, sería una forma experimental de hacer filosofía de la ciencia. En lugar de buscar modelos formales —uno de los objetivos, a veces exclusivos de la filosofía de la ciencia— desde la construcción axiomática (empirismo lógico) al predicado conjuntista (concepción estructural de las teorías científicas), nos encontraríamos con una reconstrucción empírica, *a posteriori* del proceso, basado en un procedimiento experimental de ensayo y error.

Entre los objetivos teórico-prácticos del aprendizaje automático podríamos destacar (siguiendo Korb 2001) los siguientes: ordenar un conjunto de datos no estructurados; desarrollar algún tipo de estrategia inductiva o abductiva para crear clasificaciones y descubrir regularidades; buscar la evidencia mínima necesaria para poder establecer regularidades y construir clasificaciones; y comprobar la fiabilidad predictiva que tiene el modelo obtenido.

Como podemos ver, no son objetivos tan alejados de la filosofía de la ciencia. Paul Thagard tiene una intuición similar cuando afirma que «[l]as ramas de la filosofía interesadas en el razonamiento están en continuidad con la psicología y la inteligencia artificial.» (Thagard 1988, 103). La diferencia principal radicaría seguramente en el tipo de temas en que estas disciplinas están interesadas: así mientras el eje central del aprendizaje automático es establecer modelos matemáticos aplicados de procesos inductivos y abductivos, la filosofía de la ciencia tiene otros muchos intereses como el realismo, la relación entre ciencia, técnica y sociedad, etc.

Aunque una parte relevante de la investigación en aprendizaje automático es de carácter matemático, consistente sobre todo en mejorar las técnicas existentes de análisis abductivo-estadístico y crear nuevas técnicas matemáticas, la mayor parte de sus investigadores, tal y como recogíamos en la definición de El Naga y Murhpy, se ocupan de desarrollar aplicaciones prácticas. Buena parte de la economía digital actual sería imposible sin algoritmos de aprendizaje automático que desarrollan clasificaciones automáticas o encuentran regularidades para generar predicciones. Pensemos en el algoritmo PageRank detrás del motor de búsqueda de Google, el algoritmo responsable de calcular el precio de un viaje en Uber a partir de los datos de oferta y demanda, el tiempo atmosférico, el código que hace posible que Amazon nos recomiende libros o productos electrónicos a partir de nuestras compras anteriores y lo que otros usuarios han adquirido, las listas de recomendaciones musicales de Spotify, etc. Nos costaría encontrar ahora mismo un producto exitoso de Silicon Valley que no se construya desde algoritmos, y según nuestra vida se va digitalizando progresivamente, estos algoritmos tendrán cada día más importancia.

Estos problemas son de carácter no determinista: no hay —o no se conocen— leyes deterministas que rijan ese proceso, de manera que los resultados son siempre probabilísticos, asignando probabilidades a las diferentes predicciones en función de la fiabilidad del algoritmo, que a su vez depende de la fiabilidad y consistencia de los datos. De hecho, buena parte de esos problemas no son simplemente no deterministas, sino que entran en lo que

Nigel Cross (2006) llama «wicked problems», es decir, problemas que ya no están bien definidos de salida y que se transforman según el investigador va desarrollando una solución al problema.

Es importante destacar que estos resultados no se fundamentan a partir de una demostración matemática, sino a través de una serie de procesos de ensayo y error, en los que los algoritmos son modificados, los ejemplos ampliados o reducidos en función de las predicciones generadas por el algoritmo. Aumentar la sofisticación matemática no implica necesariamente una gran mejora en resultados. Una regresión lineal pura y dura nos puede dar tranquilamente una fiabilidad del 90% y un algoritmo sofisticado de Deep learning puede acabar ofreciéndonos, después de días de ajustes y pruebas, un 92% que a efectos prácticos puede resultar completamente idénticos, Cary O’Neill y Rachel Schutt (2013).

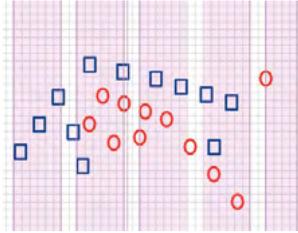
Esos modelos más complejos, basados en redes neuronales de varias capas, como toda la familia del Deep learning, son verdaderas cajas negras, en las que el investigador o investigadora no tienen ninguna posibilidad de explicar lo que el algoritmo hace de una forma comprensible. Más allá de llevar a cabo un «cluster análisis» e indicar qué «neuronas» parecen responder más a ciertos patrones de los datos y cuáles a otros, pero sin poder establecer qué tipo de computación específica realizan. Así, este tipo de modelos tienen una utilidad práctica clara, pero no nos ofrecen nuevo conocimiento que podamos explicar, pues el algoritmo no es racionalizable.

El modelo de Arthur descrito en la sección anterior se ajusta aquí perfectamente, al encontrarse una serie de necesidades con unos principios lógico-matemáticos que pueden solventar esa necesidad. Larry Page y Sergei Brin observaron la poca efectividad de los buscadores pre-Google y decidieron mejorarlos; Garrett Camp y Travis Kalanick —los fundadores de Uber— vieron una oportunidad de negocio clara modulando las tarifas de precios de un viaje en taxi en función de diversas variables que el modelo tradicional de taxis no recoge, o Jeff Bezos de Amazon capaz de mejorar considerablemente el comercio on-line al disponer de la capacidad de predecir los intereses de los usuarios.

El aprendizaje automático implica una cooperación entre modelo teórico y tinkering práctico, el equivalente informático a experimentar sin tener una idea clara, sin un modelo teórico de lo que va a salir. Para entender esta peculiar interacción entre teoría y práctica necesitamos explicar el concepto de sobreajuste (overfitting).

Un algoritmo de aprendizaje automático se basa en una serie de ejemplos para inducir una regularidad. El número de ejemplos utilizados es una magnitud crítica. Si utilizamos un número demasiado pequeño el programa no será capaz de encontrar nada relevante, pero si ponemos demasiados el algoritmo puede acabar simplemente «memorizando» los diferentes ejemplos y no extraer ninguna regla. Consideremos por ejemplo los algoritmos basados en árboles de decisión, un algoritmo estándar cuando tenemos que dividir un conjunto de ejemplos en más de dos categorías. Imaginemos que queremos predecir si una persona saldrá a pasear en bicicleta en función del tiempo. La primera variable es el viento. Si sopla demasiado viento no saldrá. La temperatura es la segunda variable. Si el viento es flojo y hace frío no saldrá. Si hace demasiado calor tampoco saldrá, si hay una temperatura intermedia sí saldrá.

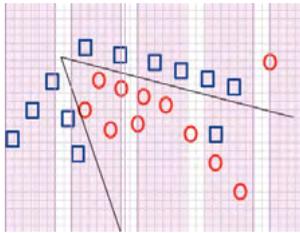
El algoritmo de árboles de decisión rápidamente encontraría la línea divisoria sin problemas y podría predecir con muy pocos errores si esa persona saldrá o no en bicicleta en función de la velocidad del viento y la temperatura. Ahora imaginemos que los datos que tenemos no son tan claros, sino que se distribuyen de forma menos ordenada, como aquí:



Esquema 1

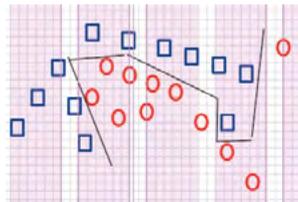
Los círculos representan días en los que la persona en cuestión no sale en bicicleta, los rectángulos en los que sí. El eje de las abscisas representa el viento y las ordenadas la temperatura.

Con un número correcto de ejemplos, el algoritmo será capaz de encontrar una división suficientemente razonable, aunque no perfecta, para poder predecir correctamente la mayoría de las veces. Gráficamente sería algo así:



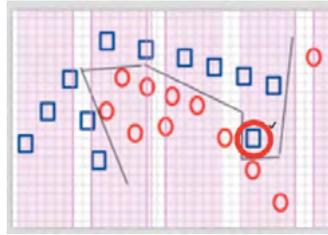
Esquema 2

Sin embargo, si ofrecemos demasiados ejemplos al programa lo que hará es desarrollar una fórmula demasiado detallada, totalmente adaptada a los ejemplos concretos, y que será incapaz de crear una generalización útil. Algo así:



Esquema 3

¿Qué implica el cambio? Básicamente, que el programa no habrá hecho una auténtica generalización y simplemente se adaptará a los ejemplos usados. Consideremos el cuadrado marcado con un círculo rojo en el esquema 4.



Esquema 4

Ese punto será clasificado como un día en el que no se puede salir en bicicleta (círculo) por el algoritmo asociado al gráfico del esquema 2, pero como un rectángulo por el algoritmo del tercer esquema. Quizás ese rectángulo en realidad es una entrada incorrecta en la base de datos, pero la diferencia más relevante aquí es que el algoritmo del esquema 2 ha aprendido una regla, mientras que el del esquema tres simplemente ha memorizado los valores de los puntos concretos.

Aunque un algoritmo basado en árboles de decisión es un constructo matemático, pura teoría, su aplicación implica un grado alto de experimentación. Para no caer en sobreajuste, hay que jugar con diversos parámetros, como el número total de ejemplos, pero también con elementos de la estructura interna del algoritmo, como el número mínimo de divisiones del árbol. Es decir, en qué momento la división del árbol en sub-ramas se detiene y damos el problema por solucionado. Como antes, si ponemos un número mínimo demasiado alto nos arriesgamos a no encontrar ninguna regularidad. Si lo hacemos muy pequeño nos arriesgamos a hacer algo similar al esquema 3: una modelización de los datos tan cercana al ejemplo concreto que no contiene ninguna generalización interesante.

Vemos así como no se trata de una aplicación deductiva del modelo teórico al caso práctico, sino de un ir y volver de ejemplo a la teoría, experimentando con diversos valores hasta que conseguimos una modelización suficientemente funcional como para ser de utilidad a la hora de recomendar grupos musicales, predecir futuros movimientos de la bolsa o establecer el precio de una carrera de taxi un sábado de madrugada en que llueve profusamente. Estos ejemplos también muestran la centralidad del factor utilidad, apuntado por los autores referidos anteriormente (Nickles 2003; Sintonen 2009; Rensulli 2003).

La elección del modelo teórico es, en buena parte, intuitiva. No hay un metamodelo teórico claro que establezca qué algoritmo es mejor para un problema dado. Frente a una necesidad como establecer si una operación con tarjeta de crédito es legal o ilegal solo el contexto y una serie de pruebas ensayo/error nos permitirán establecer si utilizar un algoritmo basado en árboles de decisión, un clasificador bayesiano ingenuo, el modelo de aprendizaje profundo (deep learning) de Google o el algoritmo SVM (Support Vector Machines) es el más adecuado. En esta decisión se combinan hábitos de trabajo, análisis matemático del problema, intuición del programador experimentado, el desarrollo de algún tipo de heurística del analista y hasta la recomendación de un antiguo profesor o un compañero de trabajo.

El desfase entre invención e innovación descrito por Arthur es manifiesto en el desarrollo del aprendizaje automático. No se trata ni siquiera de un simple lapso temporal de

esperar un cierto tiempo hasta que la invención devenga finalmente disruptiva. Es un proceso no lineal con diversos lapsus temporales y bucles. La primera generación de aprendizaje automático se basaba en el desarrollo de simples redes neuronales de una capa, que se consideraron revolucionarias hasta que Marvin Minsky les proporcionó un golpe mortal desde la teoría matemática al demostrar que no podrían solucionar algo tan sencillo como el problema de la *o* inclusiva. 30 años más tarde revivirán con la aparición de redes neuronales con una capa extra y el PDP o Procesamiento Distribuido En Paralelo. Durante más de una década se desarrollaron todo tipo de modelos innovadores basados en redes neuronales, hasta que fueron substituidas por otros sistemas de aprendizaje automático basados en la búsqueda de regularidades estadísticas. Y así las redes neuronales pierden su poder innovador a principios del siglo XXI hasta que hace unos pocos años Google las vuelve a poner de actualidad con el desarrollo de su modelo de aprendizaje profundo, responsable entre otras cosas de uno de los giales de la Inteligencia Artificial: ganar al Go a un gran maestro humano.

4. *Conclusión*

Cuando se observan las prácticas científicas en detalle resulta manifiesto que teoría y experimento tienen cada una su función y que de ninguna manera puede considerarse a la experimentación como un mero apoyo confirmador o refutador a un modelo teórico.

Cuando analizamos la forma en que las ciencias innovan en métodos y procesos epistémicos, podemos observar un fenómeno similar. El aprendizaje automático, una rama supuestamente teórica de las ciencias de la computación, contiene toda una serie de bucles extraños en los que modelo teórico matemático y experimentación práctica se van combinando entre sí sin que ninguno tenga la última palabra a la hora de establecer una innovación. Con nuestra investigación hemos querido mostrar que incluso una rama tan, en principio, teórica y apriorística —básicamente matemática aplicada—, necesita también de un fuerte componente de *tinkering* experimental. Teoría y práctica necesitan la una de la otra, ya sea en procesos teóricos o aplicados. La innovación en ciencias no es una excepción.

Agradecimiento

Este trabajo se enmarca en el proyecto del Ministerio de Ciencia e Innovación de España «Creatividad, revoluciones e innovación en los procesos de cambio científico» (referencia FFI2014-52214-P); y en el Grupo de investigación Grupo de Estudios Humanísticos sobre Ciencia y Tecnología (GEHUCT), reconocido y financiado por la Generalitat de Catalunya (referencia SGR 2014).

REFERENCIAS

- Arthur, W. Brian. 2007. The structure of invention. *Research Policy* 36: 275-287.
- Artigas, Mariano. 1989. *Filosofía de la ciencia experimental; Introducción*. Pamplona: EUNSA.
- Carayannis, Elias, Edgar González y John Wetter. 2003. The nature and dynamics of discontinuous and disruptive innovations from a learning and knowledge management perspective. En Larisa V. Shavinina, ed., *The international handbook on innovation*, 115-138. Amsterdam: Elsevier Science Ltd.
- Cross, Nigel. 2006. *Designerly ways of knowing*. London: Springer.
- Edgerton, David. 2013. *Quoi de neuf? Du rôle des techniques dans l'histoire globale*. Paris: Éditions du Seuil.
- El Naqa, Isam y Martin J. Murphy. 2015. What Is machine learning? En Isam El Naqa y Martin J. Murphy, eds. *In machine learning in radiation oncology*, 3-11. Cham: Springer International Publishing.
- Estany, Anna. 2007. Innovación tecnológica y tradiciones experimentales. Una perspectiva cognitiva. *Revista Ciencias UNAM* 88: 34-45.
- Estany, Anna y Rosa M. Herrera. 2016. *Innovación en el Saber Teórico y Práctico*. Londres: College Publications.
- Galison, Peter. 1987. *How experiments end*. Chicago: University of Chicago Press.
- Georgsdottir, Asta S., Todd I. Lubart y Isaac Getz (2003). The role of flexibility in innovation. En Larisa V. Shavinina, ed., *The international handbook on innovation*, 180-190. Amsterdam: Elsevier Science Ltd.
- Gooding, David, Trevor Pinch y Simon Schaffer. 1989. *The uses of experiment*. Cambridge: Cambridge University Press.
- Hacking, Ian. 1996. *Representar e intervenir*. México: Editorial Paidós/ UNAM.
- Iglesias, Mercedes. 2004. El giro hacia la práctica en filosofía de la ciencia: una nueva perspectiva de la actividad experimental. *Opción. Revista de Ciencias Humanas y Sociales* 20 (44): 98-119.
- Korb, Kevin B. 2001. Machine learning as philosophy of science. En Kevin B. Korb y Hilan Bensusan, eds. *Proceedings of the ECML-PKDD-01 workshop on machine learning as experimental philosophy of science*. Freiburg.
- Korb, Kevin B. 2004. Introduction: Machine learning as philosophy of science. *Minds and Machines* 14(4): 433-440.
- Marinova, Dora y John Phillimore. 2003. Models of innovation. En Larisa V. Shavinina, ed., *The international handbook on innovation*, 44-53. Amsterdam: Elsevier Science Ltd.
- Martínez, Sergio M. 2003. *Geografía de las prácticas científicas*. México: UNAM.
- Mitchell, Tom M. 1997. *Machine learning*. Redmond, WA: McGraw-Hill.
- Nickles, Thomas. 2003. Evolutionary models of innovation and the Meno problem. En Larisa V. Shavinina, ed., *The international handbook on innovation*, 54-78. Amsterdam: Elsevier Science Ltd.
- O'Neill, Cary y Rachel Schutt. 2013. *Doing data science*. Boston: O'Reilly.
- Ordoñez, Javier y José Ferreirós. 2002. Hacia una filosofía de la experimentación. *Critica. Revista Hispanoamericana de Filosofía* 34/102:47-86.
- Pickering, Andrew. 1995. *The mangle of practice. Time, agency and science*. Chicago: The University of Chicago Press.
- Renzulli, Joseph S. 2003. Three-ring conception of giftedness: Its implications for understanding the nature of innovation. En Larisa V. Shavinina, ed., *The international handbook on innovation*, 79-96. Amsterdam: Elsevier Science Ltd.
- Sintonen, Matti. 2009. Tradition and innovation: exploring and transforming conceptual structures. En Joke Meheus y Thomas Nickles (eds.) *Models of discovery and creativity*, 209-221. Dordrecht: Springer.
- Stan, Jérémy. 2012. *Les 100 inventions les plus marquantes*. Paris: Editions ESI.
- Sundbo, Jon. 2003. Innovation and strategic reflexivity. En Larisa V. Shavinina, ed., *The international handbook on innovation*, 97-114. Amsterdam: Elsevier Science Ltd.
- Thagard, Paul. 1988. *Computational philosophy of science*. Cambridge, MA: MIT Press/Bradford Books.

DAVID CASACUBERTA es profesor contratado doctor de Filosofía de la Ciencia en la Universidad Autónoma de Barcelona. Su línea de investigación actual se centra en analizar cómo las tecnologías digitales están transformando los procesos cognitivos, tanto a nivel epistemológico como sus implicaciones culturales y sociales. Recibió el premio Eusebi Colomer de la Fundación Epson al mejor ensayo sobre los aspectos sociales, antropológicos, filosóficos o éticos relacionados con la nueva sociedad tecnológica con su libro *Creación colectiva*. También es ganador del premio Ingenio 400 organizado por el Ministerio de Cultura y la Sociedad Estatal de conmemoraciones culturales al mejor proyecto de net.art con su obra (en colaboración con Marco Bellonzi) X-Reloaded. Ha trabajado para Trànsit Projectes (www.transit.es) en diversos proyectos financiados por la Unión Europea relacionados con la e-inclusión y las TIC y en cómo pueden utilizarse para luchar contra la exclusión social. En paralelo ha trabajado en proyectos de cultura digital como Autobahn —una cartografía de la música electrónica para la Mediateca del Caixaforum e Hiperiment— el hipertexto en juego para el Centro de Cultura Contemporánea.

DIRECCIÓN: Departamento de Filosofía, Edificio B, Campus de la UAB (08193 Bellaterra-Cerdanyola del Vallès). España. Email: david.casacuberta@uab.cat

ANNA ESTANY es catedrática de Filosofía de la Ciencia en la Universidad Autónoma de Barcelona. Ha sido visiting scholar en University of California San Diego, en la École Normale Supérieure-París y en el Institut d'Histoire et de Philosophie des Sciences et des Techniques de París. Sus líneas de investigación son modelos de cambio científico, enfoque cognitivo en filosofía de la ciencia y de la tecnología y filosofía de las ciencias de diseño. Es autora de *Modelos de cambio científico* (1990), *Vida, muerte y resurrección de la conciencia* (1999), *La fascinación por el saber* (2001), *¿Eureka? el trasfondo de un descubrimiento sobre el cáncer y la genética molecular* (2003) junto con D. Casacuberta, *Innovación en el saber teórico y práctico* (2016) junto con Rosa M. Herrera, y coordinadora junto a Ángel Puyol de *Filosofía de la epidemiología social*.

DIRECCIÓN: Departamento de Filosofía, Edificio B, Campus de la UAB (08193 Bellaterra-Cerdanyola del Vallès). España. E-mail: anna.estany@uab.cat