

Sobre las limitaciones de Big Data en ciencias sociales

On the limits of Big Data in social sciences

IGNACIO C. MAESTRO CANO

Centro Asociado Francisco Tomás y Valiente
Universidad Nacional de Educación a Distancia
C/ Casa de la Misericordia 34. 46014 Valencia

icmaestro@valencia.uned.es

ORCID: <https://orcid.org/0000-0001-5082-2086>

Este artículo está sujeto a una: "[Creative Commons Reconocimiento-No Comercial](#)" (CC-BY-NC).

DOI 10.24197/st.2.2019.77-98

RECIBIDO: 22/12/2018

ACEPTADO: 21/03/2019

Resumen: Big Data ha irrumpido como una prometedora infraestructura tecnológica, garantía de grandes progresos en ciencias sociales. La recogida masiva de información sobre hábitos y características de los usuarios de dispositivos móviles se propone como clave para una descripción precisa de la sociedad y sus individuos. No obstante, el entusiasmo ante esta nueva fuente de datos descubierta y su posibilidad de interconexión (redes sociales, compras online, etc.) podría suponer una nada desdeñable amenaza para un adecuado progreso científico, que diera origen a cierto neoqueteteletismo tecnológico que confunda la *datificación* del mundo con su comprensión. Surge así el riesgo de transferir la atención a los datos, desatendiéndose contexto y metodología. Big Data puede suponer una importante contribución, pero una mala comprensión podría tornarla engañosa.

Palabras clave: Big Data, redes sociales, técnicas cuantitativas, epistemología, sociología.

Abstract: Big Data has burst as a promising technological infrastructure, as a guarantee for great advances in social sciences. The massive data collection about habits and characteristics of mobile device users has been suggested as the key to reach an accurate description of the society and its individuals. However, the enthusiasm concerning this new data source and its interconnection capability (social networks, online shopping, etc.) could suppose a non-insignificant threat to a suitable scientific progress, raising some kind of technological Neoqueteteletism, and making the world's "datification" be mistaken for world's comprehension. So, there is a risk of transferring the attention to the data alone ignoring context and methodology. Big Data can be a remarkable contribution; however, any misunderstanding may be seriously misleading.

Keywords: Big Data, social networks, quantitative techniques, epistemology, sociology.

1. INTRODUCCIÓN

Una de las ideas que subyace tras lo que se ha dado en llamar Big Data es que, en la actualidad, los dispositivos móviles constituyen una fuente de información sin precedentes acerca de cuestiones de lo más diversas relativas a los usuarios de tales dispositivos. Se ha afirmado que los ciudadanos están evolucionando “de actores urbanos a indicadores urbanos” (Martino et al., 2010: 1). Más aún, la proliferación de dispositivos móviles de todo tipo (*smartphones*, *tablets*, GPS, entrenadores inteligentes, etc.) ha dado origen a lo que se ha denominado (Wolf, 2010) la autocuantificación o el “yo cuantificado” (*quantified self*, en inglés). La tesis propuesta es que la recopilación de todo tipo de datos relativos a la propia persona y su entorno, fundamentalmente para el seguimiento y control de la salud y los hábitos personales (alimentación, ejercicio, sueño, estados de ánimo...), ha de posibilitar una casi perfecta caracterización de cada individuo. En el caso de este yo “cuantificado”, se trata de introducir los números “en los acogedores confines de la vida personal” (Wolf, 2010), sin embargo estos “grandes datos” están introduciéndose en muchos otros ámbitos de nuestra vida: salud, desplazamientos, compras, opiniones, etc. Es la llamada Revolución Big Data.

Si a esta facilidad en la obtención de datos se le añade la capacidad de conectar unas fuentes de información con otras (cuentas de correo, videos reproducidos, compras online, comentarios en Twitter, etc.), resulta comprensible que surja cierta expectación en torno a las posibilidades que abre el acceso a dicha información. Dicho esto, es aquí donde debieran terciar la prudencia y el sentido común. Una primera objeción que podría plantearse es que no es lo mismo ser un usuario de un teléfono inteligente que ser una persona y que se está excluyendo a cierto segmento de la población, el cual variará según la edad, la zona geográfica, etc. Big Data podría introducir por tanto un sesgo de partida.

Como se ha adelantado, Big Data está siendo presentado como una revolución y, además, como una revolución científica, pese a que por el momento no pasa de presentarse como una reedición más del acostumbrado positivismo, si acaso ahora con ropajes más *geek*; en la forma de una descomunal base de datos. El planteamiento viene a ser: conozco mucho al ser humano porque tengo muchos datos sobre él. Pero lo cierto es que, por poner un ejemplo, cada tuit recabado sobre un individuo no incrementa necesariamente la información disponible ya que muchas veces se trata de mensajes retuiteados. Se aprecia pues un primer escollo y es que la obsesión por recopilar cualquier dato, va a generar gran cantidad de ruido e información redundante. El conocimiento no es un mero acumular información. En este sentido, la pretensión de Big Data frente al método científico tradicional (Anderson, 2008) se presenta, cuando menos, equívoca.

Si desea estudiarse la movilidad en determinada área metropolitana y se dispone de información relativa a los desplazamientos de los usuarios de un *smartphone* (que, como se ha dicho, no son todas las personas), es posible volcar esos millares de coordenadas y datos temporales en una base de datos y empezar a buscar la “aguja en el pajar”. Pueden explorarse patrones de movimiento o preferencias. Aquellos lugares en los que la población parece poseer cierta “preferencia” por estar en determinado momento: un polígono industrial, una instalación deportiva o un centro comercial. Lo que habría que determinar es si las conclusiones así inferidas no habrían sido posibles reuniendo a tres o cuatro expertos que conozcan mínimamente la región y utilicen sus conocimientos y el sentido común. ¿Se precisan millones de datos para saber que los niños entran y salen de los colegios a determinadas horas? ¿Los necesitamos para saber que los fines de semana los centros comerciales reciben la mayoría de sus visitas semanales? ¿Es necesario “curiosear” en los dispositivos de la población —más o menos con su consentimiento— para alcanzar este tipo de progreso? Puede que en ocasiones sea así pero, en muchas otras, es posible que esto no sea sino una distracción sobre lo verdaderamente importante. Así, en el mejor de los casos, Big Data ofrecería la respuesta a una infinidad de preguntas indiscriminadas que (quizás por algún motivo) jamás fueron formuladas desde un ámbito científico, cuando lo que a menudo se necesitan son unas pocas preguntas, pero bien escogidas.

No tiene sentido centrar la atención sobre los datos sin considerar el contexto o la propia experiencia. Como especialistas, es necesario abordar cada campo de estudio de manera orgánica. Por poner un símil tecnológico muy oportuno, con Big Data podríamos terminar haciendo caso únicamente a la información de nuestro navegador GPS (los datos), sin poner la debida atención a la carretera (al contexto). En ese conjunto inabarcable de datos que constituyen los Big Data, a menudo resulta imposible obtener una visión global; esa necesaria contextualización que en ocasiones le resulta en cambio tan fácil de adquirir a un experto sin necesidad de acceder a tanta información.

Por supuesto que deben seguirse aplicando toda clase de técnicas cuantitativas, y que para éstas la recopilación de la suficiente información resulta crucial, pero no debiera prescindirse de la prudencia que impone la complejidad e impredecibilidad de las cuestiones que a menudo han de abordarse. No sabemos con certeza si nuestros modelos y pronósticos nos aportan siempre la información correcta, pero siempre —y ese puede ser el problema— nos aportarán confianza. Al fin y al cabo, ese es el propósito de cualquier modelo: permitir que nos movamos con cierta confianza a través de determinado problema. Pues bien, se trata de conservar esta confianza como un bien y no convertirla en una de las principales amenazas que puedan cernirse sobre el progreso científico.

2. BIG DATA. ¿MODELIZANDO EL CAOS?

No son pocos los especialistas que han señalado la dificultad inherente existente en la tarea de abordar de un modo científico aquello que atañe al ser humano. Desde las propias ciencias sociales, se ha advertido de la indiscutible diferencia existente entre éstas y las ciencias naturales. Es algo que ya en el siglo XVI Montaigne (1965) expresaría de muy diversas maneras: “los actos de los hombres (...) parece imposible que hayan salido del mismo cosechero [*boutique*]” (p. 15); “No tengo nada que decir de mí, de forma concluyente, simple y sólida, sin confusión y sin mezcla, ni en una palabra. Distingo es el elemento más universal de mi lógica” (p. 20); “Se encuentra tanta diferencia entre nosotros y nosotros mismos, como la que existe entre nosotros y los demás hombres” (p. 22). Por su parte, y ya en el siglo XX, Hayek (1978) advirtió de la “esencial complejidad” de los fenómenos a estudiar por las ciencias sociales. Al optar por el término “esencial”, Hayek deja entrever que no se trata de una diferencia de grado sino en la naturaleza de ésta, de modo que, a priori, dicha complejidad no podría vencerse mediante una mera acumulación de datos. Para Hayek, “los aspectos de los sucesos a explicar y de los cuales podemos obtener datos cuantitativos son necesariamente limitados y pueden no incluir los importantes”. Más aún, alerta de que “en ciencias sociales, a menudo se trata como importante lo que [simplemente] resulta ser susceptible de ser medido”.

En consonancia con tales apreciaciones, y en el marco de sus estudios de *soft psychology*, Meehl censuró la ligereza con la que se presentaban los resultados de los contrastes de hipótesis en psicología, que en ocasiones alcanzaban unos sensacionales niveles de confianza del 0,001%. En contraste con estos resultados, Meehl (1978: 807) no dudaba en afirmar que bastarían “10 minutos de reflexión superficial” para encontrar “20 rasgos que hacen a la psicología humana difícil de abordar de manera científica [*hard to scientize*]” y concluir que quizás “debiéramos reconocer la posibilidad de que nunca vaya a existir una teoría realmente impactante [*impressive*] en la psicología social o de la personalidad. No me agrada pensarlo, pero sencillamente podría ser cierto” (p. 829). Y es que, siguiendo lo ya expresado por Lykken (1968), Meehl (1997) recuerda que “siendo la estructura causal y compositiva del pensamiento y la sociedad (...) la que es, casi todas las variables que medimos están correlacionadas en alguna medida” (p. 393) de modo que “en las ciencias sociales todo se correlaciona con casi todo lo demás” (p. 402).

Fue fundamentalmente a partir del siglo XIX cuando las expectativas puestas en el progreso científico y las técnicas estadísticas en lo relativo al conocimiento de la sociedad y el comportamiento de sus individuos comenzaron a crecer de un modo apasionado. Así, se afirmaba:

“No sólo en lo relativo a sus cualidades físicas, sino también en lo que se refiere a sus acciones, el hombre se encuentra sometido a la influencia de causas que son en su mayoría regulares y periódicas; y tienen efectos igualmente regulares y periódicos. Es posible, a través de un estudio ordenado, determinar las causas y su forma de actuar o las leyes que las dan origen” (Quetelet, 1835a: 13).

Hoy se afirma que es posible, recogiendo “las migas de pan digitales que todos dejamos atrás a medida que avanzamos por el mundo —registros de llamadas, transacciones con tarjeta de crédito y posiciones GPS—”, predecir quien es más probable que “padezca diabetes” (Pentland, 2014: 16) o “si alguien es la clase de persona que devolverá los créditos” (p. 7), algo que permitirá “construir una sociedad que sea mejor a la hora de evitar la crisis de mercados, la violencia étnica y religiosa, los estancamientos políticos, la propagación de la corrupción y las peligrosas concentraciones de poder” (p. 16). Se está claramente ante aquella “pretensión del conocimiento” de la que advirtió Hayek (1978) cuando se lamentara de “la tendencia a imitar lo más de cerca posible los procedimientos de las brillantemente exitosas ciencias físicas —un intento que en nuestro campo [la economía] puede conducir a un rotundo error—”.

En el ámbito de la economía, ya Knight (1921), uno de los fundadores de la Escuela de Chicago, abordó la cuestión de la inevitable presencia de incertidumbre. Para ello, quiso distinguir entre una incertidumbre medible (aquella para la cual disponemos, por ejemplo, de una distribución de probabilidades) y otra no medible. Sin embargo, lejos de sentirse paralizado por dicha realidad, Knight vino a argumentar que, en realidad, es prácticamente gracias a esta incertidumbre que resulta posible hablar de una “inteligencia” en la gestión, de una auténtica toma de decisiones: “Con la ausencia de la incertidumbre (...) resulta dudoso si la inteligencia misma existiría” (p. 268). Con dicha apreciación, lo que Knight hizo en el fondo fue destacar el papel del administrador, esto es, del individuo, en la economía y, con ello, situándole en un plano claramente por encima del que pueda concederse a la información. Frente a una época que ensalza la avalancha de datos (*data deluge*) en la que estamos inmersos y prácticamente la hace administradora del progreso científico, Knight admite y destaca —casi podría decirse que celebra— la presencia de incertidumbre.

En las decisiones del ser humano intervienen muchas variables (o a veces pocas, pero de naturaleza desconocida), lo que ocasiona que estas se tornen impredecibles. Para empezar, debieran tenerse presentes los que se conocen como sesgos cognitivos, muy poderosos pero difíciles de diagnosticar. A este respecto, resulta interesante consultar los resultados publicados desde 1979 por Makridakis y Hibon. Entre otras cosas, descubrieron que “los métodos estadísticamente más sofisticados o complejos no proporcionaron pronósticos más exactos que los más sencillos” (Makridakis y Hibon, 2000: 453), algo que

podría entenderse como una advertencia para aquellos que consideran que todo es una cuestión de recabar más datos. Big Data —disponer de muchos más datos— puede que nos permita disponer de un cuadro más preciso de la realidad de hoy, pero no nos permite pronosticar mucho más acerca del cuadro que tendremos mañana. No siempre más significa mejor.

Pero pronosticar seduce. Tanto si se acierta como si no. Quizás por ello, tal y como también advirtieron Makridakis y Hibon (2000: 454), esta “fuerte evidencia empírica ha sido ignorada por los estadísticos teóricos (...). En su lugar, han concentrado sus esfuerzos en construir modelos más sofisticados sin considerar la capacidad de tales modelos de predecir datos de la vida real con más exactitud”. De este modo, se han ensayado todo tipo de modelos para tratar de explicar la existencia de determinados patrones o leyes en la evolución de la situación social. Por ejemplo, se pretende desentrañar procesos como la pérdida o persistencia de los rasgos culturales de las distintas comunidades en el marco de la globalización gracias al uso de filtros adaptativos (Axelrod, 1997: 203), pero lo cierto es que, incluso cuando semejante modelo pudiera quedar validado por algún caso real concreto, difícilmente pueda admitirse que modelos de este tipo posean una validez general. Y sin una validez general, no tiene mucho sentido hablar de “ley” o modelo alguno.

Puede aducirse que un problema es el de la pretensión de identificar modelos para las cuestiones propias de las ciencias sociales y otro distinto es el de la cantidad de información disponible o requerida para determinarlos. Sin embargo, lo que desde el punto de vista de la epistemología o la filosofía de la ciencia se aprecia tras ambas cuestiones es un mismo problema: el de las pretensiones de los distintos campos científicos. Es algo de lo que ya han advertido algunos autores, como es el caso de Boyd y Crawford (2012), quienes citando un trabajo de Venoit (2007) han recordado:

“En la era del giro computacional, resulta cada vez más importante reconocer el valor de los «datos pequeños». (...) Su trabajo [el de Venoit] cuenta una historia que no podría ser descubierta mediante el análisis [*farming*] de millones de cuentas de Facebook o Twitter y contribuye a su campo de investigación de forma significativa, a pesar de contar con una mínima cantidad de participantes (...): *en algunos casos, menos es mejor*” (Boyd y Crawford, 2012: 670) (la cursiva ha sido añadida).

Por otro lado, las capacidades de las técnicas Big Data no se verán notablemente alteradas, al menos en lo que se refiere a las limitaciones aquí analizadas, por una alteración o mejora en lo relativo al procesamiento de los datos (evidentemente sí que afectará a la sencillez de su implementación o a su mejor adecuación al problema en cuestión que se esté tratando). De este modo, independientemente de cuál sea la implementación de estas técnicas, por ejemplo, a través de la elección de un lenguaje de programación u otro (como es bien sabido, en este contexto destaca el empleo de los lenguajes Python y R), el

resultado será, en lo fundamental, muy similar. Sea con R, con Python o con otro lenguaje, la inexistencia de un modelo estadístico fiable que pueda asociarse al problema abordado (el comportamiento humano) permanecerá. Se trata aquí, en definitiva, de limitaciones que tienen que ver más con el fenómeno analizado —epistemológicas, podría decirse—, que con las herramientas utilizadas para ello.

Así, la debilidad de la información disponible, e incluso de las técnicas empleadas, es algo en constante debate. Un problema que es posible apreciar incluso en la sencilla regresión lineal, una técnica estadística que ofrece siempre, y de un modo que casi se podría calificar de “sumiso”, unos resultados sobre los que recae en no poca medida la enorme responsabilidad de que a las ciencias sociales se las tome por tales. Por supuesto que es posible aplicar complejos modelos ARMA o filtros de Kalman al comportamiento de un vehículo cuyo peso, empuje y características aerodinámicas son conocidos, y el cual se ve exclusivamente sometido a las leyes de la dinámica. Pero este no es el caso de un individuo, mucho menos el de una agrupación de ellos. No tiene sentido aplicar en tal caso unas hipotéticas leyes de la “dinámica social” sobre individuos cuyas tendencias, inercias y aceleraciones (circunstancias, preferencias o *élan* vital) son poco conocidas. Por mucho que se disponga de ingentes cantidades de información al respecto. El ser humano no es en absoluto equiparable a ningún cuerpo inerte, por mucho que pueda haber situaciones en las que su comportamiento pueda mostrarse suficientemente estable.

3. LOS DATOS NO HACEN A LA CIENCIA

Una de las primeras conclusiones que se desprenden de lo expuesto hasta aquí es que una cosa es disponer de muchos datos acerca de un fenómeno y otra muy distinta es conocerlo bien. Ello no hace sino resaltar la diferencia existente entre el pensar y el mero computar. El mero acumular datos no es ciencia, ni lo será jamás. No se trata sólo de “cuántos” sino de “cómo” y de “cuáles”. Han de buscarse por tanto *Smart Data* y no *Big Data*. De este modo, la cantidad de datos no es equiparable con la cantidad de “verdad”, siendo aplicable aquí aquella frase de Ortega y Gasset (1914: 18): “¿sabemos tantas cosas que no comprendemos!”.

Volviendo al ejemplo ya mencionado, es posible el empleo de distintos algoritmos para conocer de forma aproximada los patrones de movilidad en determinada área metropolitana y así poder abordar la necesaria planificación territorial. Ello es indispensable y, puestos a acometer dicha tarea, debería hacerse del modo que se entienda más racional. No obstante, este proceder no garantiza que la realidad vaya a ajustarse demasiado al modelo obtenido. Dicho modelo constituye todo lo que podemos y debemos hacer, pero no ha de perderse de vista que no es la realidad, sino tan sólo una posibilidad más o

menos racional, más o menos exacta. Recurriendo al conocido paralelismo: “un mapa no es el territorio” (Korzybski, 1933: 750).

No se trata de proponer una supuesta imposibilidad de llevar a cabo predicciones suficientemente precisas en ciencias sociales, sino de señalar que tales predicciones deberían restringirse con frecuencia a las cuestiones más generales o globales. Quizás fuera conveniente prescindir de la pretensión de realizar predicciones sobre problemas demasiado concretos o específicos ya que, como señalara Hayek (1978), “actuar en la creencia de que poseemos el conocimiento y el poder que nos permite moldear los procesos de la sociedad completamente a nuestro gusto, un conocimiento que en realidad no poseemos, es posible que nos cause mucho daño”. Al fin y al cabo, un reconocimiento de las naturales limitaciones existentes en las ciencias sociales —al igual que en cualquier otra disciplina— no supone menosprecio alguno.

Lo cierto es que la inmensa mayoría de los modelos actualmente utilizados en economía y sociología operan en un marco utilitarista conocido como teoría de la elección racional y que se remonta a Edgeworth (1881). De acuerdo con este marco, los individuos toman sus decisiones disponiendo de un conocimiento perfecto de las distintas alternativas y buscando siempre y exclusivamente maximizar su propio beneficio. Se trata por tanto de un modelo que asume un comportamiento *perfectamente* egoísta y *perfectamente* racional: el llamado *homo aeconomicus*. Como era previsible (el mismo concepto de “racional” no está exento de problemática), no han faltado los críticos a este planteamiento (Sen, 1989; Elster, 1993; Green y Shapiro, 1994), de los que ha surgido lo que se ha dado en llamar economía post-autista. Así, Sen considera que resulta excesivamente simplista ignorar otros intereses, como puedan ser todos aquellos situados a medio camino entre el interés propio y el de todos (los intereses de grupo, por ejemplo). De un modo tan sintético como oportuno, Sen defiende que “el egoísmo universal como *realidad* puede ser falso, pero el egoísmo universal como requisito de la *racionalidad* es evidentemente absurdo” (Sen, 1989: 33). De hecho, en su opinión, los principios de semejante *homo aeconomicus* no son sino los de un verdadero “imbécil social”, un tonto sin sentimientos que no pasa de ser un sujeto irreal sin compromiso, ni simpatía, ni valores: “el hombre puramente económico es casi un retrasado mental desde el punto de vista social” (Sen, 1986: 202). Partiendo de críticas como esta se han desarrollado diversos modelos alternativos como el de racionalidad limitada (Simon, 1957). Si bien estos modelos siguen encorsetando la realidad, proponen un planteamiento más “heurístico” que “algorítmico” y optan por sugerir una solución no plenamente justificada, antes que por calcular un resultado que haya de adoptarse como indiscutiblemente correcto.

En definitiva, puede afirmarse que el ser humano es un ser racional, pero “racional” no significa predecible. Del mismo modo en que “más racional” no significa “más humano”. Es más, en línea con las apreciaciones de Sen, casi

podría decirse que si el ser humano fuese un ser *perfectamente* —esto es, exclusivamente— racional, dejaría de ser humano. Todo ser humano tiene sentimientos e intuiciones (su propio y particular modo de percibir la realidad), lo que hace que en ocasiones se comporte de forma irracional. No se trata sólo de que sea poco probable encontrar un comportamiento humano que sea siempre sea racional, sino que se ha llegado a hablar de la “«irracionalidad» fundamental [inherente a la idea] de una actitud ante la vida que fuese perfectamente «racional»” (Knight, 1921: 62). Tampoco debiera pasarse por alto que, este planteamiento acerca de una supuesta predictibilidad del comportamiento humano, se asienta sobre una cuestión tan fundamental como es la del libre albedrío. Si somos capaces de predecir con total precisión el comportamiento de un individuo, ¿dónde queda su libertad? Es una cuestión que, al margen de sus evidentes connotaciones religiosas (Venn, 1866: 339), ya le fue criticada en el siglo XIX a Quetelet y otros “físicos sociales” (Venn, 1866: 343-370). En este contexto, el propio Venn recuerda la distinción entre causalidad y regularidad, un concepto con una clara conexión con el incremento de la cantidad de datos disponibles y, por ende, con el fenómeno Big Data: “la mera regularidad de la estadística observada (...) me temo que apenas tiene alguna conexión con la causalidad” (p. 336). Sucede, en cierto modo, lo expresado por Taleb (2008: 251), para quien “la ubicuidad de la campana de Gauss no es una propiedad del mundo, sino un problema de nuestra mente que surge del modo en que contemplamos aquél”. Es el *It from bit* que, a modo de variante del llamado principio antrópico, ha propuesto Wheeler (1990: 310). Y según el cual, toda variable (cada *it*) tiene su origen en la propia observación que de él hace el hombre (el *bit* o la información recogida por el ser humano).

También Venn (1866: 344) había llamado la atención sobre la posible interferencia que, al modo de una variante social del principio de incertidumbre de Heisenberg, la mera observación de los acontecimientos pudiera ocasionar sobre el propio fenómeno estudiado. Se trata en definitiva de recordar que nos encontramos en un universo participativo: “cuando las cosas cuyas estadísticas estamos debatiendo son acciones que realizan hombres exactamente iguales a nosotros y que podemos llevar a cabo algún día” (p. 353), “cuando la inferencia es acerca de la conducta de seres humanos, a menudo se olvida que con la inferencia en sí, si es publicada, podemos haber producido una fuente insospechada de perturbación” (p. 345). Un caso paradigmático de dicho fenómeno podrían constituirlo por ejemplo los resultados de los sondeos electorales a través del denominado efecto arrastre o *bandwagon*. Para Venn, la conclusión es clara:

“Podemos realizar las afirmaciones y críticas que deseemos acerca de la conducta *pasada* [énfasis en el original] de los hombres, pero directamente nos comprometemos con cualquier afirmación relativa al futuro (...) No pretenderán, por

supuesto, ser capaces de vaticinar las acciones particulares de los individuos, pero constantemente reivindican que es bastante posible que algún día [¿con el advenimiento de Big Data?] seamos capaces de vaticinar tendencias generales, así como las consecuencias de la conducta de grandes masas de hombres” (Venn, 1866: 345-346).

Y para ilustrar el problema, Venn cita a John Stuart Mill, para quien la situación de las ciencias sociales era análoga a aquella por la que pasó la astronomía en su momento, que, siendo inicialmente una ciencia no exacta, terminó por serlo:

“La ciencia de la naturaleza humana es de este tipo. Dista mucho de los estándares de exactitud ya alcanzados en astronomía; pero no hay ninguna razón por la que no pueda ser tan ciencia como lo es la mareografía, o como lo fue la astronomía cuando sus cálculos sólo habían dominado los principales fenómenos, no así sus perturbaciones” (Mill, 1843: 493).

Ciertamente, lo aquí descrito no se trata de una problemática que, en un sentido estricto, tenga su origen exclusivamente en las técnicas Big Data, pero sí posee una clara conexión con éstas. Una conexión que se manifiesta en la interpretación del papel desempeñado por los datos en la fiabilidad y exactitud de las conclusiones. Podría así llegarse al extremo ya advertido por algunos autores:

“La ciencia dura de la física, con sus medidas precisas y leyes mecánicas exactas, se ha convertido en una metáfora, un modelo de racionalidad y disciplina que la gente trató de imitar conforme estudiaban materias más suaves tales como la biología, el lenguaje, el comportamiento humano o incluso el de sociedades enteras. De algún modo parecía no importar que ninguna de estas materias se prestara a medidas y predicciones precisas. La ciencia había dejado de ser simplemente una metodología y se había convertido también en una ideología” (Rampton y Stauber, 2002: 37).

4. PRÉSTAMOS CIENTÍFICOS Y MODELIZACIÓN SOCIAL

Se ha recordado la importancia —la necesidad más bien— de adoptar por una actitud cautelosa con relación a las perspectivas derivadas del notable aumento en la cantidad de información disponible que supone Big Data. Ello es resultado de la relativa proliferación de todo tipo de expectativas que describen el panorama Big Data como si de un auténtico Shangri-La para las ciencias sociales se tratara. Por citar algún ejemplo en esta dirección, Ball (2012) ha defendido que, con independencia de su “aparente” complejidad, la explicación y predicción del comportamiento social no están fuera del alcance de la ciencia. En su opinión, se trataría simplemente de sustituir “átomos y moléculas por

personas, coches, operadores de mercados o empresas” (Ball, 2003), para alcanzar aquella “física social” de la que ya en el siglo XIX, hablara Auguste Comte (1830-1842: 22). De este modo, y de acuerdo con Tuhtan (2013: 68), “al emplear la física estadística, el comportamiento de un sistema de muchos cuerpos no se considera desde el punto de vista del individuo, sino que es visto más bien como la culminación de los inevitables efectos en la acción colectiva”. Tomando así a los seres humanos por partículas dentro de una especie de laboratorio social, determinados físicos, “hambrientos, adrenalizados” y “tomando prestadas ideas y técnicas de cualquier sitio”, según lo ha expresado Watts (2013), auguran el comienzo de “un periodo apasionante y de grandes descubrimientos”. Surgen de este modo, con relativa asiduidad, propuestas que plantean dilemas tales como: “¿La física simplemente nos ayuda a explicar y comprender, o podemos emplearla [también] para anticipar y de este modo evitar problemas, mejorar nuestras sociedades, hacer un mundo mejor y más seguro?” (Ball, 2004: 7). Quizás queriendo evitar sobresaltos, se ha visto la necesidad de aclarar que, pese a todo, “existen diferencias físicas fundamentales a la hora de considerar los sistemas que contienen personas o partículas [físicas] que van más allá de lo trivial” (Tuhtan, 2013: 73) o que “la diferencia entre una persona y una partícula es fundamentalmente de perspectiva. Idealmente no hay ninguna diferencia, pero sí la hay en la realidad” (p. 74). Resultaría admisible ignorar tales afirmaciones si no fuera porque con ellas se están generando unas expectativas perniciosas por cuanto, por su origen, revisten de una autoridad científica resultados que, en su mayoría, no pasan de ser discretos, si no meras conjeturas. Baste para ilustrar la cuestión el mencionar cómo se ha llegado a afirmar que “en el futuro —y antes de lo que pudiéramos pensar— (...) [sería posible] incluso identificar «criminales» antes de que en efecto uno cometa un crimen” (Mayer-Schönberger y Cukier, 2013: 12).

Siempre habrá personas a las que este escenario le parezca de lo más idílico pero la realidad es terca y entraña que tratar la información relativa a la conducta del ser humano del mismo modo que la relativa a los cuerpos celestes, la atmósfera o un compuesto químico se antoje disparatado. Pese a ello, se recurre con relativa facilidad a las más sugerentes teorías, tales como la teoría del caos o la dinámica “auto-organizada” de sistemas no-lineales, los fractales, los autómatas celulares, etc. “Es posible describir los ecosistemas utilizando las leyes de la termodinámica”, se nos dice (Tuhtan, 2013: 74). Si, es posible, pero... ¿Es acertado? ¿Práctico? En principio, cualquier préstamo científico o transvase terminológico entre ciencias puede resultar enriquecedor, además de atractivo: energía libre de Gibbs como indicador de cambio social (Tuhtan, 2013: 72), maduración de Ostwald, entalpía “social”... Sin embargo, éste debe estar justificado, evitando excesos que podrían acercarnos más a una especie de “escolástica” sociológica —en el sentido etimológico más original del término *Skholè*: “ocio”— o sociología “recreativa” que a una auténtica sociología. Por

otra parte, todo símil es eso, sólo un paralelismo. Nunca una equivalencia. Por ello, salvo que quiera caerse en un relativismo epistémico o una mera “impostura intelectual” (Sokal y Bricmont, 1998), las analogías sirven tradicionalmente para explicar un fenómeno a través de otro más accesible o sencillo, y no al revés. La sociología debe seguir siendo capaz de distinguir los hechos de aquello que, en el contexto de la econometría, Armstrong (1978) denominó el “folklore”.

Tras esta predisposición a acudir a metodologías desarrolladas por otras ciencias y diseñadas para problemas muy distintos no se oculta sino el llamado problema de demarcación. Lo único que se consigue con tal proceder es reavivar la controversia en torno a la inclusión o no de las ciencias sociales (sociología, economía, etc.) dentro de un apartado especial de ciencias “blandas” (*soft sciences*) frente a las ciencias denominadas “duras” (*hard sciences*) como son la física, la química o la medicina. Es en este contexto de una pretensión de incluir a las ciencias sociales en un hipotético apartado de ciencias “duras” en el que surge Big Data. Como si de la mera adición de más datos a una ciencia blanda fuera a surgir, de manera automática, una ciencia dura: *Sociología + Big Data = Física social*. Pero, evidentemente, esta ecuación no es correcta. Es por ello que se ha afirmado: “Números, números, números (...) La sociología ha estado obsesionada con la meta de convertirse en una ciencia cuantitativa” (Latour, 2010: 145). Y lo cierto es que, al margen de la conveniencia o no de establecer dos categorías de ciencia —una más “estricta” o “dura” y otra más “blanda”—, suele argumentarse que, mientras en las llamadas ciencias “duras” el progreso se fundamenta en el llamado consenso o paradigma científico (Kuhn, 1962), algo que garantiza que su contenido ha sido plenamente aceptado por toda la comunidad científica (no se da el caso de que haya una “escuela antigravitacional” que rechace la mecánica de Newton, por poner un ejemplo), esto no sucede así en el caso, por ejemplo, de las distintas teorías económicas.

Ya Horkheimer (1947) abordó la cuestión en el marco de su revisión de las consecuencias que, sobre el progreso científico, pudiera tener el pragmatismo predominante en la ciencia moderna y, en concreto, la centralidad de la experimentación (la obsesión por los datos):

“Es más bien el profesor de filosofía el que, esforzándose por imitar al físico para incluir su disciplina entre las «ciencias que avanzan con éxito», trata los pensamientos como si fueran cosas y elimina cualquier idea de verdad que no se corresponda con aquella que resulta del dominio racionalizado [*streamlined*] de la naturaleza” (p. 50).

Es en este contexto en el que, dado que el objeto de estudio de Big Data es el ser humano, esta tecnología ha tenido una buena acogida por parte de las ciencias sociales. Si ya el desarrollo de la estadística supuso un importante

apoyo para las ciencias sociales, resulta comprensible que se tenga por beneficioso el tratamiento de “grandes datos”. Si las técnicas cuantitativas han sido de gran ayuda, mayor parece que será ésta si la cuantía de datos empleados va en aumento. Tal y como ya ha sido expresado, se hace posible caer en el razonamiento de que “más datos es mejor”. A este respecto se ha afirmado:

“Big Data ofrece a las disciplinas humanísticas una nueva forma de reivindicar el estatus de ciencia cuantitativa y método objetivo. Hace cuantificables a muchos más espacios sociales [pero]. En realidad, trabajar con Big Data sigue siendo subjetivo y lo que cuantifica no puede reclamar necesariamente un derecho más sólido con respecto a la verdad objetiva —especialmente cuando se trata de mensajes procedentes de las redes sociales” (Boyd y Crawford, 2012: 666-667).

Cuando un investigador se enfrenta al estudio de un fenómeno que, por su propia naturaleza (i.e., ineludiblemente), manifiesta un comportamiento que podría calificarse de indómito, no tiene sentido forzar la realidad imponiéndole unos límites artificiales, sea con la adopción de un determinado modelo probabilístico o el uso de alguna función de utilidad. Tampoco la recogida de más datos alterará la naturaleza del fenómeno que los genera, al contrario, su impredecibilidad se mantendrá intacta. En este sentido, el comportamiento humano posee tal cantidad de grados de libertad que excede de todo lo que la imaginación pueda siquiera concebir. En este caso, la “textura” de la realidad es siempre muy superior a la de cualquier modelo matemático que se nos ocurra, por más complejo y aproximado que lo pretendamos concebir y por muy sugerente que se muestre: algoritmo memético (Moscato, 1989), modelo autorregresivo integrado de media móvil (Box y Jenkins 1976), modelo autorregresivo con heterocedasticidad condicional (Engle, 1982), etc. A menudo la realidad es que las posibilidades de tales herramientas estadísticas, invocadas en el campo de las ciencias sociales no son comparables a las que sí son habituales, por ejemplo, en la ingeniería.

5. ESTUDIO DE UN CASO. OBAMA Y EL “MILAGRO” BIG DATA

Uno de los más destacados asuntos en que se anunció “a bombo y platillo” la capacidad de Big Data para la explicación y casi vaticinio del resultado de un acontecimiento sociológico fueron las elecciones presidenciales de 2012 en Estados Unidos. En este caso se trató del empleo de información disponible en las redes sociales para identificar, movilizar o persuadir a electorado a favor del candidato demócrata, Barack Obama. Se trataba de aplicar lo que se conoce como *microtargeting*. Si Big Data fue realmente decisivo en los resultados electorales es algo que, por más que se argumente o ambicione, no es posible saber. De hecho, si fuera tan evidente la influencia de las técnicas utilizadas entonces, éstas habrían sido aplicadas de nuevo para

decidir el resultado de las elecciones de 2016. Pero a algunos, la realidad no les va a estropear una buena historia y, para expresarlo de manera escueta y exenta de toda exageración, se desató en los medios la “fiebre” Big Data. Baste para ilustrarlo lo expresado apenas un mes más tarde por Issenberg (2012): “Detrás de todo eso hubo puntuaciones describiendo votantes concretos: una nueva divisa política que predijo el comportamiento de sujetos humanos. La campaña no solo sabía quién eras; supo *exactamente* cómo convertirte en el tipo de persona que quería que fueses” (la cursiva ha sido añadida).

Con estos antecedentes, en las elecciones de 2016 (en las que ambos candidatos contaron con sendos equipos de *Data Analytics*) fueron numerosas las crónicas que explicaron cómo Big Data ya sabía que el triunfo sería de Donald Trump. Eso sí, lo hicieron por lo general... pasadas las elecciones. Se acudió en apoyo de Big Data con afirmaciones como: “son los humanos los que fallan, no los datos” o “tan sólo necesitamos llegar a ser más creativos [sea lo que sea lo que esto signifique] en la observación de éstos” (Timms, 2016). Pero lo cierto es que todo apuntaba a que aquellos que consideraron que era una cuestión de mera adquisición de datos, habían errado. No cabe duda de que, con ello, Big Data queda en entredicho, tal y como han defendido diversos autores (Piatetsky-Shapiro, 2016; Associate Press, 2016; Lohr y Singernov, 2016; Laurent, 2016). Se había anunciado (Crovitz, 2016) que “los datos de las pasadas elecciones presidenciales [las de 2012] descubren una innegable correlación: el candidato con los mejores datos es el ganador”. Pero al final gana el candidato que consigue más apoyo y no el que consigue más datos. Después de todo, tal y como muy oportunamente se ha recordado, “la elección de los votantes es secreta y sólo puede medirse mediante votaciones” (Sides y Vavreck, 2014).

No cabe duda de que definir una estrategia en base a lo que se recoge en 15 terabytes de datos puede dar mejores resultados que si se confía exclusivamente en el simple olfato del politólogo o sociólogo de turno, por muy inspirado que éste sea. Intuición y teoría están bien, pero resultan estériles si no se dispone de información. Tal y como ha reconocido Hurwitz (2012):

¿Serán los grandes datos la diferencia entre el éxito y el fracaso de una campaña política? La respuesta es sí y no. Si la analítica de los grandes datos es empleada estratégicamente para respaldar un plan, el éxito posee una gran oportunidad. Si los grandes datos son la estrategia, se tiene una receta para el fracaso.

En resumen, el empleo de una mayor cantidad de datos no constituye ninguna garantía de éxito, ni es llave de una hipotética verdad sociológica. Para unos, (Wolf, 2010), “estamos obligados a guiarnos por conjeturas. Seguimos a nuestro estómago. Quiero decir, algunos de nosotros lo hacemos. Otros [los que tienen éxito, se sobrentiende] utilizan datos”. Para otros (Sides y Vavreck, 2014), todo esto “suena a grandes números (...), «métricas de vanidad»:

números que parecen impresionantes pero no revelan mucho sobre la eficacia. La pregunta crucial no es cuántos contactos se hicieron, sino cuántos contactos produjeron realmente el resultado deseado: un voto para Obama”. De hecho, la propia Amelia Showalter, directora de la firma que desarrolló las técnicas Big Data de la campaña de Obama, ha reconocido: “sólo porque tengamos millones de «compartir» no significa que vayan a transformarse en millones de votos (...) No tengo ni idea de si el programa de intercambio dirigido [*targeted sharing*] nos reportó 100 ó 100.000 votos adicionales” (Sides y Vavreck, 2014). De acuerdo con la estimación de Sides y Vavreck (2014), “si los resultados de un estudio de 2010 de Facebook son algún indicio (...), el programa de la campaña movilizó sólo a 20.000 votantes —una ínfima parte del margen de cinco millones de votos de Obama”.

No se trata de menospreciar la utilidad de las nuevas tecnologías y las redes sociales en el ámbito de los estudios sociológicos, sean de carácter político o de otro tipo. Ésta está fuera de toda duda, y las posibilidades descriptivas de las redes sociales como “contenedores” de opinión pública están ahí. Sin embargo, otra cuestión es su capacidad predictiva y/o prescriptiva, esto es, la capacidad de influir de manera decisiva sobre determinadas decisiones de los individuos. Ni los intereses comerciales o mercantilistas, ni el afán de notoriedad debieran entremezclarse con aquellos intereses que le son propios a la ciencia. Desde la sociología, como ciencia, y como ciencia interesada precisamente en la sociedad, habría de velarse con especial empeño por que la aureola de respetabilidad que confiere toda técnica de investigación cuantitativa no termine por dañar el prestigio alcanzado tras décadas de recto y estoico avance.

6. SÍNTESIS CONCLUSIVA

Como se ha visto a lo largo de este trabajo, con Big Data se reclama una nueva capacidad científica pero, en gran medida y si se examina con atención, la realidad no va, por el momento, mucho más allá de la mera retórica. Sea que la *datificación* del mundo ha dado a luz a Big Data o que Big Data ha devenido en un nuevo escenario de *datificación* del mundo, se aprecia una notable falta de capacidad crítica frente a tal fenómeno y un dudoso progreso tecnológico puede abrirse paso mientras el individuo no puede sino encogerse de hombros y plegarse ante él. En ciencia, la lógica habitual hasta la fecha ha dictado que, cuanto mayores sean las promesas de una nueva técnica o teoría (más “falsable” se presenta, por utilizar la terminología de Popper), mayores han de ser los controles a los que sea sometida. Especialmente cuando, tal y como ya se ha advertido desde la revista Nature (2007: 637-638), “cualquier dato relativo a sujetos humanos plantea inevitablemente problemas de privacidad y los riesgos reales de abusar de tales datos son difíciles de cuantificar”. Es por ello que

resulta esencial poner a la ciencia a salvo de lo que podría denominarse meras tendencias o modas, evitando que inciertas promesas (como pudiera ser el caso de Big Data) echen por tierra la credibilidad de los métodos cuantitativos en ciencias sociales.

Se ha señalado la importancia de no obviar que Big Data sólo tiene en consideración a aquellos individuos que son más “habituales” en las redes sociales y que éstos, aún siendo una porción considerable, no son todos. La sociología no debiera olvidar a aquellos cuya “huella digital” sea ínfima. Es algo sobre lo que Crawford (2013) insistió a propósito del análisis de los datos recogidos en las redes sociales con motivo del paso por Nueva York (año 2012) del huracán Sandy: “Se presupone que los datos reflejan de manera exacta el mundo social, pero existen lagunas significativas, con poca o ninguna señal procedente de determinadas comunidades”. Las relaciones sociales, del tipo que sean, son lo que cuenta para las ciencias sociales y, en este sentido, la propia Crawford ha advertido de que “conforme nos desplazamos hacia una era en la que los dispositivos personales son vistos como *proxies* de las necesidades públicas, corremos el riesgo de que las desigualdades ya existentes se hagan más profundas”.

Por otra parte, se ha alertado de una serie de consecuencias derivadas de un uso inadecuado de las nuevas tecnologías y que, pese a no estar directamente relacionados con el propósito de este trabajo, sí debieran ser tenidos en consideración a la hora de valorar el papel de las técnicas Big Data en el progreso de las ciencias sociales. Se ha hablado, por ejemplo, de un cambio en la actitud tradicional hacia ciertos valores y en concreto de cierta insensibilización con respecto a la violencia o la intransigencia, así como del riesgo de despersonalización en un contexto tecnológico en el que comienza a alertarse de una cierta “desconexión entre el mundo real y el virtual” (López, 2016); del problema que surge cuando la separación entre mundo real y virtual se hace tan borrosa que deja de percibirse. Parece cumplirse así el pronóstico situacionista de Debord (1967) y, en la sociedad tecnológica actual, “todo lo que era vivido directamente ha derivado en una representación” (párrafo 1) en la que se aprecia una “degradación del ser en tener (...) un desplazamiento generalizado del tener al aparentar” y en la cual “el sujeto no puede surgir más que de la sociedad [de las redes sociales, se diría hoy]” (párrafo 52) y donde los medios de comunicación desempeñan un papel clave como intermediarios. Es el espectáculo —ahora en su versión 2.0— como “instrumento de unificación” (párrafo 3). Tal y como se ha expresado, siendo esta cuestión distinta a la aquí tratada —las capacidades de Big Data en el ámbito de la sociología—, tiene una clara conexión con ésta por cuanto es en este contexto de relaciones (que distorsiona la realidad de un individuo o incluso la del conjunto de la sociedad) en el que Big Data recaba una información que toma como representativa o digna de toda confianza.

Se trata en definitiva de evitar caer en un nuevo queteletismo, o cualquier otra forma postmoderna de determinismo sociológico, aunque este se oculte bajo capas y más capas de datos. Aquel *homme moyen* de Quetelet (1835a y 1835b), que “resumiría sobre sí mismo a un época dada” (Quetelet, 1835b: 276) y sobre el que se intenta establecer “las bases de una *física social*” (Quetelet, 1835a: 21), en realidad, difícilmente describa la realidad del individuo, menos aún la de la sociedad.

Para sus defensores, Big Data lo pone todo alcance de la mano en el terreno científico pero, como se ha descrito aquí, a pesar de la utilidad de la información recabada y la virtud que supone su recogida en grandes cantidades, lo cierto es que, en el peor de los casos, Big Data bien pudiera terminar por mostrarse una falacia en el ámbito científico, una burbuja en el empresarial y un riesgo en el privado. Antes que una creativa explosión de conocimiento, Big Data podría devenir una estéril y devastadora inundación de datos.

No es ninguna forma de escepticismo lo que aquí se propugna. Se trata más bien de una cautelosa búsqueda de rigor. Lo aquí expresado busca tan sólo estimular una oportuna reflexión en torno a las verdaderas capacidades de lo que se ha dado en llamar Big Data. De hecho, es posible que la mejor forma de preservar su innegable valor sea evitar los diversos sensacionalismos aparecidos en los últimos años, algunos de los cuales han sido reseñados aquí. Si hablamos de demoscopia o *marketing*, no cabe duda que el conocimiento de determinadas características o hábitos de los potenciales votantes o clientes resulta de interés. Pero si hablamos de hacer ciencia, la cosa cambia. Conocer es mucho más que recabar datos y conocer al ser humano o la sociedad, suele resultar mucho más complejo que persuadir o vender.

En lo fundamental, y por muchos que sean los datos recabados, el ser humano sigue siendo el mismo: imperfecto, impredecible y de razón limitada. Seguimos siendo, a fin de cuentas, “hijos del azar y la adversidad” (Nietzsche, 1954: 30).

REFERENCIAS BIBLIOGRÁFICAS.

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine*, 16 julio de 2008 (en línea). http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory, último acceso 23 de octubre de 2017.

Armstrong, J.S. (1978). Forecasting with Econometric Methods: Folklore Versus Fact. *Journal of Business*, 51(4), 549-564.

https://repository.upenn.edu/cgi/viewcontent.cgi?article=1008&context=marketing_papers

- Associate Press (2016). Donald Trump Wins 2016 US Election: Why Big Data Is Amongst the Biggest Losers in US Elections. *Gadgets 360*, 11 de noviembre de 2016 (en línea). <http://gadgets.ndtv.com/others/features/why-big-data-is-amongst-the-biggest-losers-in-us-elections-1624244>, último acceso 23 de octubre de 2017.
- Axelrod, R. (1997). The dissemination of culture. A model with local convergence and global polarization. *Journal of Conflict Resolution*, 41(2), 203-226. <https://doi.org/10.1177/0022002797041002001>
- Ball, P. (2003). *The Physics of Society*. Conferencia en la London School of Economics, 6 de marzo de 2003.
- Ball, P. (2004). *Critical Mass: How One Thing Leads to Another*. London: Heinemann.
- Ball, P. (2012). *Why Society is a Complex Matter. Meeting Twenty-first Century Challenges with a New Kind of Science*. Berlin/Heidelberg: Springer.
- Box, G.E.P. y Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Boyd, D. y Crawford, K. (2012). Critical Questions For Big Data. Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662-679.
- Comte, A. (1830-1842). *Cours de philosophie positive. Tome Premier*. Paris: Rouen Frères.
- Crawford, K. (2013). The hidden biases in big data. *Harvard Business Review* (en línea). <https://hbr.org/2013/04/the-hidden-biases-in-big-data>, último acceso 23 de octubre de 2017.
- Crovitz, L.G. (2016). Trump's Big Data Gamble. *The Wall Street Journal*, 24 de Julio de 2016 (en línea). <http://www.wsj.com/articles/trumps-big-data-gamble-1469395312>, último acceso 23 de octubre de 2017
- Debord, G. (1967). *La Société du spectacle*. Paris: Buchet-Chastel.

- Edgeworth, F.Y. (1881). *Mathematical Psychics. An Essay on the Application of Mathematics to Moral Sciences*. London: C. Kegan Paul & Co.
- Elster, J. (1993). Some unresolved problems in the theory of rational behaviour. *Acta Sociologica*, 36(3), 179-189. <https://doi.org/10.1177/000169939303600303>
- Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4), 987-1007. <https://doi.org/10.2307/1912773>
- Green, D.P. y Shapiro, I. (1994). *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science*. New Haven: Yale University Press.
- Hayek, F.A. von. (1978). La Pretensión del Conocimiento (conferencia en homenaje de Alfred Nobel, pronunciada el 11 de diciembre de 1974). En *Los Premios Nobel de Economía 1969-1977* (pp. 245-258). México: Fondo de Cultura Económica.
- Horkheimer, M. (1947). *Eclipse of Reason*. New York: Oxford University Press.
- Hurwitz, J. (2012). The Making of a (Big Data) President. *Business Week*. Bloomberg, 14 de noviembre de 2012 (en línea). <http://www.bloomberg.com/bw/articles/2012-11-14/the-making-of-a-big-data-president>, último acceso 23 de octubre de 2017.
- Issenberg, S. (2012). How President Obama's campaign used big data to rally individual voters. *MIT Technology Review*, 12 de diciembre de 2012 (en línea). <http://www.technologyreview.com/featuredstory/509026/how-obamas-team-used-big-data-to-rally-voters>, último acceso 23 de octubre de 2017.
- Knight, F.H. (1921). *Risk, Uncertainty and Profit*. New York: Augustus M. Kelley.
- Korzybski, A.H.S. 1933. *Science and Sanity. An Introduction to Non-Aristotelian Systems and General Semantics*. New York: Institute of General Semantics.

- Kuhn, T.S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Latour, B. (2010). Tarde's Idea of Quantification. En M. Candea (ed.) *The Social After Gabriel Tarde: Debates and Assessments* (pp. 145-162). London: Routledge.
- Laurent, S. (2016). L'élection de Trump et les trois échecs du "big data" electoral. *Le Monde*, 10 de noviembre de 2016 (en línea). http://www.lemonde.fr/les-decodeurs/article/2016/11/10/l-election-de-trump-et-les-trois-echecs-du-big-data-electoral_5028978_4355770.html, último acceso 23 de octubre de 2017.
- Lohr, S. y Singernov, N. (2016). How Data Failed Us in Calling an Election. *The New York Times*, 10 de noviembre de 2016 (en línea). <http://www.nytimes.com/2016/11/10/technology/the-data-said-clinton-would-win-why-you-shouldnt-have-believed-it.html>, último acceso 23 de octubre de 2017.
- López Pérez, P.M. (2016). EnREDados, del uso al abuso de las Redes Sociales Virtuales. *XII Congreso Español de Sociología. Grandes transformaciones sociales, nuevos desafíos para la sociología*. 30 de junio, 1 y 2 de julio de 2016.
- Lykken, D.T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, 60(3), 151-159.
- Makridakis, S. y Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, 16, 451-476. [http://dx.doi.org/10.1016/S0169-2070\(00\)00057-1](http://dx.doi.org/10.1016/S0169-2070(00)00057-1)
- Martino, M., Britter, R., Outram, C., Zacharias, C., Biderman, A., y Ratti, C. (2010). *Senseable city*. Cambridge: MIT Senseable City Lab.
- Mayer-Schönberger, V. y Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. New York: Houghton Mifflin Harcourt.
- Meehl, P.E. (1978). Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology. *Journal of Consulting and Clinical Psychology*, 46(4), 806-834. <http://dx.doi.org/10.1037/0022-006X.46.4.806>

- Meehl, P.E. (1997). The problem is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. En L.L. Harlow, S.A. Mulaik y J.H. Steiger (eds.), *What if there were no significance tests?* (pp. 393-425). Mahwah: Erlbaum.
- Mill, J.S. (1843). *A System of Logic, Ratiocinative and Inductive. Vol. II.* London: John W. Parker.
- Moscato, P. 1989. On Evolution, Search, Optimization, Genetic Algorithms and Martial Arts: Towards Memetic Algorithms. *Caltech Concurrent Computation Program. Report 826.* Pasadena: California Institute of Technology.
- Montaigne, M. de. (1965) [1595]. *Essais. Livre second.* Paris: Gallimard.
- Nature (2007). A matter of trust. *Nature*, 449.
- Nietzsche, F. (1954). Die Geburt der Tragödie oder Griechentum und Pessimismus. En K. Schlechta (ed.), *Friedrich Nietzsche: Werke in drei Bänden*, Band 1. München: Carl Hanser.
- Ortega y Gasset, J. 1914. *Meditaciones del Quijote.* Madrid: Revista de Occidente.
- Pentland, A.S. (2014). *Social physics: How good ideas spread—The lessons from a new science.* New York: Penguin Books.
- Piatetsky-Shapiro, G. (2016). Trump, Failure of Prediction, and Lessons for Data Scientists. *KDnuggets* (en línea). <http://www.kdnuggets.com/2016/11/trump-shows-limits-prediction.html>, último acceso 23 de octubre de 2017.
- Quetelet, L.A.J. (1835a). *Sur l'homme et le développement de ses facultés, ou Essai de physique sociale. Tome Premier.* Paris: Bachelier.
- Quetelet, L.A.J. (1835b). *Sur l'homme et le développement de ses facultés, ou Essai de physique sociale. Tome Second.* Paris: Bachelier.

- Rampton, S. y Stauber, J. (2002). *Trust Us, We're Experts PA: How Industry Manipulates Science and Gambles with Your Future*. New York: Penguin Putnam.
- Sen, A. (1986). Los tontos racionales: una crítica de los fundamentos conductistas de la teoría económica. En F. Hahn y M. Hollis (eds.), *Filosofía y teoría económica* (pp. 172-217). México: Fondo de Cultura Económica.
- Sen, A. (1989). *Sobre ética y economía*. Madrid: Alianza.
- Sides, J. y Vavreck, L. (2014). Inconclusive Results: Obama's Not-So-Big Data. *Pacific Standard*, 21 de enero de 2014 (en línea). <http://www.psmag.com/books-and-culture/obamas-big-data-inconclusive-results-political-campaigns-72687>, último acceso 23 de octubre de 2017.
- Simon, H.A. (1957). *Models of Man: Social and Rational*. New York: John Wiley and Sons.
- Taleb, N.N. (2008). *The Black Swan. The Impact of the Highly Improbable*. New York: Random House.
- Timms, A. (2016). Is Donald Trump's Surprise Win a Failure of Big Data? Not Really. *Fortune*, 14 de noviembre de 2016 (en línea). <http://fortune.com/2016/11/14/donald-trump-big-data-polls>, último acceso 23 de octubre de 2017.
- Tuhtan, J.A. (2013). On the Differences Between a Person and a Particle. *Journal of Human Thermodynamics*, 9(5), 65-76.
- Veinot, T. (2007). The Eyes of the Power Company: Workplace Information Practices of a Vault Inspector. *The Library Quarterly*, 77(2), 157-180. <https://doi.org/10.1086/517842>
- Venn, J. (1866). *The Logic of Chance: An Essay on the Foundations and Province of the Theory of Probability, with Especial Reference to Its Application to Moral and Social Science*. London/Cambridge: Macmillan and Co.
- Watts, D. (2003). *Six degrees: The science of a connected age*. New York: Norton.

Wheeler, J.A. (1990). Information, physics, quantum: The search for links. En W. H. Zurek (ed.), *Complexity, Entropy, and the Physics of Information* (pp. 3-28). Redwood City: Addison-Wesley.

Wolf, G. (2010). The Data-Driven Life. *The New York Times Magazine*, 28 de abril de 2010 (en línea). <http://www.nytimes.com/2010/05/02/magazine/02self-measurement-t.html>, último acceso 23 de octubre de 2017.