

FILOSOFIA ȘTIINȚEI

CERCETAREA DIRIJATĂ DE DATE ȘI REVIGORAREA ȘTIINȚEI BACONIENE ÎN „ERA PETABYȚILOR”

GABRIEL NAGĂȚ

Institutul de Filosofie și Psihologie „Constantin Rădulescu-Motru” al Academiei Române

Abstract. Is the „end of theory” really in our sight? Was scientific understanding overrated? Is the Big Data revolution revealing some unnoticed limitations of our old conception of knowledge and scientific inquiry? Are we witnessing the final days of our concept of scientific method and the arising of a „fourth paradigm” in science? These are some of the main questions that confront the philosophers of science, when faced with the Data Deluge of our days. In this article, I try to examine if, and how far, the Big Data analysis challenges our current understanding of science, by calling into question the received view of „the scientific” (i.e., hypothetico-deductive) method, but also the meaning or utility of some central epistemological concepts: context of discovery – context of justification, correlation – causation, explanation, (dis)unity of science etc. I think that, at the end of the day, the discussions about how to scientifically use the huge datasets could be seen as manifestations of a pluralistic idea of science (and a pluralistic vision in the contemporary philosophy of science), where „pluralism” should be interpreted as a sort of *virtuous relativism*, cautiously separated from the supposed *vicious relativism* of the 1970s.

Keywords: Big Data; the end of theory; theory-driven research; data-driven research; Baconian science; the fourth paradigm of science

1. Într-un articol voit provocator prin însuși titlul pe care îl purta (*The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*)¹, cunoscutul jurnalist și antreprenor Chris Anderson, coordonatorul conferințelor TED, atrăgea atenția asupra faptului că progresele contemporane în domeniul achiziționării și prelucrării informației ar determina intrarea omenirii în „Era Petabyților” (*Petabyte Age*), care s-ar deosebi de epocile anterioare prin simplul fapt că „mai mult înseamnă diferit”². „Diluviul” datelor care sunt permanent obținute în tot mai

¹ C. Anderson, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, „Wired Magazine” 16 (07), 2008. www.wired.com/science/discoveries/magazine/16-07/pb_theory.2008

² „The Petabyte Age is different because more is different” (C. Anderson, *op. cit.*). Să amintim că unitatea primară de măsură a informației este *bit*-ul, 8 biți = 1 Byte, iar în expresie zecimală 1 PB = 1000 TB = 1000² GB = 1000³ MB = 10¹⁵ B. Pentru interpretarea acestui ordin de mărime, să precizăm că în 2018 întreg materialul stocat în celebra bibliotecă americană Library of Congress însuma circa 16 PB (fără a lua în calcul și copiile diverselor documente), dar astăzi există baze de date care se măsoară în Exabyți (1 EB = 1000 PB) și ne îndreptăm deja către „Era Zettabyților” (1 ZB = 1000000 PB).

multe domenii ar reclama, printre altele, și o mutație majoră la nivelul cercetării științifice contemporane, mutație despre care se presupune că ar putea consemna chiar „sfârșitul teoriei” și al „metodei științifice”. Altfel spus, în noile circumstanțe tehnologice cercetarea științifică ar avea șansa de a se reinventa prin ieșirea din acea multiseclară fază a construcțiilor teoretice, conturată odată cu modernitatea, și prin deplasarea către o nouă modalitate de raportare la obiectele pe care ne propunem să le cunoaștem, una în care investigațiile „metodice” practicate până acum și-ar dovedi caracterul desuet, iar „metoda științifică” (în acea accepțiune pe cât de uzuală, pe-atât de vagă a expresiei) ar deveni inutilă³.

La scara Petabyților, susține Anderson, informația obținută pe diverse căi ajunge să fie privită ca o simplă chestiune de „statistică agnostică”⁴, abordarea matematică ocupând atunci primul plan în examinarea datelor, și abia după stabilirea corelațiilor matematice (în acord cu principiul „With enough data, the numbers speak for themselves”) ar apărea chestiunea identificării unui context relevant care să permită eventuale interpretări non-matematice. Atunci când datele sunt în cantități uriașe și numerele dau impresia că ar putea „vorbi de la sine”, pare tentant să se renunțe și la căutarea unor relații de cauzalitate, urmând să ne mulțumim doar cu identificarea unor simple corelații, ceea ce ar presupune acceptarea ideii că „știința poate avansa chiar și în lipsa unor modele coerente sau a unor teorii unificate”. Provocarea lui Anderson ar semnala atunci și apariția unui *nou mod de producere a cunoașterii*, așadar o preocupare pentru ceea ce filosofii obișnuiesc să numească *contextul descoperirii*, cu toată că mare parte a secolului trecut s-a considerat că singurul interes filosofic veritabil ar fi de găsit doar în chestiunile ce țin de contextul *justificării* sau întemeierii. O asemenea manieră de practicare a științei ar însemna, într-adevăr, abandonarea tradiționalei forme de raportare ipotetico-deductivă la obiectele cercetării (pe traseul ipoteză – model – testare), iar prin renunțarea la tentativele de interpretare cauzală a corelațiilor descoperite ar fi subminat și *idealul explicativ* în favoarea unui „nou mod de înțelegere a lumii”⁵.

³ „This is a world where massive amounts of data and applied mathematics replace every other tool [...]. Out with every theory of human behaviour, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves” (C. Anderson, *op. cit.*). Să remarcăm că sunt invocate aici doar situații ce țin de domeniul științelor socio-umaniste, dar în alte locuri sunt menționate și aspecte specifice științelor naturii (mecanica cuantică sau biologia, de exemplu).

⁴ O „statistică agnostică” (*agnostic statistics*) vizează cercetarea a „ceea ce poate fi aflat despre lume fără a presupune că există un model generativ simplu despre care să se știe că este adevărat”, și în acest sens ea fundamentează un tip de inferență statistică utilă acelor cercetători care se feresc să facă „assumptions beyond what they or their audience would find credible” (P. Aronow, B. Miller, *Foundations of Agnostic Statistics*, Cambridge, Cambridge University Press, 2019, p. xv).

⁵ „Petabytes allow us to say: «Correlation is enough». We can stop looking for models. We can analyse the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot. [...] Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all” (C. Anderson, *op. cit.*).

În opinia unor filosofi ai științei mai înclinați spre un tradiționalism intransigent, prin afirmațiile sale atât de tranșante Chris Anderson ar fi dovedit că s-a lăsat sedus de mirajul unei forme radicale de empirism și de un „hype-blinded high-tech optimism”⁶, ceea ce înseamnă, finalmente, că nu ar fi înțeles mare lucru din spiritul care animă metoda cercetării științifice și care îi și determină trăsăturile definitorii. De fapt, susțin acei critici, numerele nu pot să „vorbească” pur și simplu „de la sine” și, în orice caz, ar fi greu de admis ideea că atunci când am renunța la formularea unor ipoteze și la căutarea unor modele teoretice am mai putea pretinde că facem încă „știință”, câtă vreme acest termen va fi asociat cu ideea căutării de *explicații*, și nu doar cu interesul de a se identifica simple corelații sau pattern-uri greu inteligibile și non-cauzale⁷. Altfel spus, atunci când este corect înțeleasă și practică, știința contemporană se dovedește în continuare un tip de activitate în care este nevoie de „mai multe teorii și mai puține date”⁸.

De pe poziții mai tolerante, alți comentatori – cei mai mulți fiind ei înșiși cercetători implicați direct în domeniul științei datelor (*data science*, așadar *data scientists*⁹) – iau ceva mai în serios concluziile și sugestiile cu tentă filosofică ale celor preocupați de evaluarea și gestionarea problemelor concrete pe care le ridică fenomenul Big Data și, mai ales, tipul de analiză reclamat de el (*new data analytics*). Privită dintr-o asemenea perspectivă mai *open-minded*, identificarea pe cale pur matematică, algoritmică, a unor pattern-uri (adică abordarea cunoscută și sub denumirea „big data/machine learning/artificial intelligence”) pare un procedeu metodologic promițător și acceptabil, cel puțin în anumite situații, însă numai cu condiția de a fi înțeles ca un demers preliminar ce s-ar cuveni urmat, cândva, de un altul – cel al formulării unor teorii care să permită interpretarea acelor pattern-uri și

⁶ S. Succi, P. V. Coveney, *Big Data: The End of the Scientific Method?*, „Philosophical Transactions of the Royal Society A”, 377, 2019, 20180145. <http://dx.doi.org/10.1098/rsta.2018.0145>

⁷ „If we stop looking for models and hypotheses, are we still really doing science? Science, unlike advertising, is not about finding patterns – although that is certainly part of the process – it is about finding explanations for those patterns” (M. Pigliucci, *The End of Theory in Science*, EMBO Reports, 10, 6, 2009: 534).

⁸ M. Frické, *Big Data and Its Epistemology*, „Journal of the Association for Information Science and Technology”, 2014. DOI: 10.1002/asi.23212

⁹ Faptul că limba română nu și-a impus un echivalent scurt și neechivoc pentru englezescul *scientist* (*Wissenschaftler*, *scientifique*, *scienziato*, *científico*), traducătorul fiind silit să facă echilibristică cu termeni precum „cercetător” și „om de știință” (atunci când vrea să traducă, de exemplu, *natural scientist*) este, probabil, mărturisirea nesățioasei apetențe pentru fantazare și eseistică în prea „mioritică” noastră cultură modernă. Așa se face că în toate dicționarele românești „scientist” este doar „adeptul scientismului”, termen cu o conotație aproape exclusiv negativă. Citez cu nesaț, de câte ori îmi crez prilejul, niște fraze „bine simțite” ale lui A.D. Xenopol, care era de părere că „în împrejurările în cari ne aflăm” (anul 1869!) ideea progresului țării prin știință „este cu deosebire periculoasă pentru noi”, aplicarea ei în școlile românești putând să ducă la situația în care „spiritele alese” să fie atrase „către ocupațiuni științifice ca singurele prin care pot bine merita de țara lor”. Aflăm, deci, că studiul științei ar fi riscat să influențeze negativ tineretul nostru, în sensul că „un spirit care s-ar fi dedat literaturii, s-ar apuca de științele naturii”, și prin această alegere „s-ar răpi de la destinațiunea lor sau chiar cu totul națiunei noastre rarele comori ale inteligenței” (A.D. Xenopol, *Istoria civilizațiunei* [1869], în *Scrieri sociale și filosofice*, București, Editura Științifică, 1967, p. 168).

identificarea unor explicații¹⁰. Aşa cum remarcă Luciano Floridi, adevărata problemă epistemologică pe care o ridică fenomenul Big Data este cea a pattern-urilor prea neînsemnate (*small patterns*): în domenii precum genetica, neuroştiinţele sau fizica experimentală trebuie să dobândim capacitatea de a înţelege unde anume am putea să găsim acele noi pattern-uri realmente capabile să adauge valoare cercetărilor noastre, dar să învăţăm şi în ce fel ar putea fi ele exploatare pentru avansul cunoaşterii¹¹. Să reţinem, de aici, şi sugestia că impactul noilor tehnologii de strângere şi analiză a unor mari cantităţi de date ar putea fi semnificativ mai important în anumite domenii sau ştiinţe decât în altele, ceea ce confirmă, o dată în plus, naivitatea celebrei teze a unităţii ştiinţei.

Ecoul acelui articol al lui Anderson se propagă şi astăzi printr-o multitudine de comentarii generate mai ales în lumea acelor *data-scientists*, dar observaţii importante au venit şi dinspre cercetători din ştiinţele naturii şi, desigur, dinspre filosofii ştiinţei şi tehnologiei. Voi examina în continuare câteva dintre raportările cele mai semnificative la provocarea lansată de el şi voi pune în evidenţă cele două direcţii pe care s-ar înscrie proiectele de reformare a metodei ştiinţifice clasice, pe care se presupune că le-ar sugera fenomenul Big Data: (1) un *empirism extrem*, pur inductivist, care ne-ar conduce, în fond, spre o revigorare a „ştiinţei baconiene”¹² şi (2) o reformare a metodei ştiinţei în direcţia a ceea ce se numeşte *cercetare dirijată de date* (*data-driven research*), o formă de practică ştiinţifică privită uneori ca fiind complementară în raport cu *cercetarea teoretic-dirijată* (*theory-driven research*) sau, alteori, pur şi simplu ca o extensie sau dezvoltare contemporană a acesteia. Anticipând finalul unui demers care ar solicita, cu siguranţă, o sumedenie de comentarii şi clarificări, voi spune că ideile promovate de Anderson şi de alţi autori preocupaţi de impactul pe care îl are (sau ar trebui să-l aibă) tehnologia informaţiei asupra cercetării ştiinţifice se încadrează în tendinţa mai generală de promovare a unei perspective *pluraliste* în filosofia contemporană a ştiinţei, o perspectivă pe care ar impune-o chiar caracterul pluralist al ştiinţei înseşi. Ar fi însă vorba despre un pluralism ce se doreşte şi se afirmă ca un fel de *relativism virtuos*, emancipat de orice urmă a celui relativism cumva „vicios” apărut în cea de-a doua jumătate a secolului trecut şi care a cochetat, uneori, cu ideea „anarhismului” metodologic.

2. Apelând la o terminologie deja consacrată în literatura domeniului, am putea spune că mai toţi comentatorii lui Anderson consideră că aprecierile lui cu privire la impactul progreselor înregistrate în ştiinţa şi tehnologia informaţiei semnaleză o problemă reală, aceea a posibilităţii (sau, în unele cazuri, chiar a necesităţii) de a se trece de la tradiţionalul model al cercetării ştiinţifice *teoretic-*

¹⁰ S. Succi, P. V. Coveney, *op. cit.*, p. 11.

¹¹ L. Floridi, *Big Data and Their Epistemological Challenge*, „Philosophy of Technology” 25 (4), 2012, p. 436: „We need more and better techniques and technologies to see the small data patterns, but we need more and better epistemology to sift the valuable ones”.

¹² Asupra încercărilor spre care ne poate împinge folosirea acestei sintagme, voi reveni ceva mai jos.

dirijată (theory-driven) la modelul unei cercetări bazate pe *principiul prevalenței sau priorității datelor*. Orientarea cercetării pornind de la acest principiu poate primi însă două înțelesuri distincte – ba chiar *radical* deosebite, așa cum se va vedea în continuare.

Din perspectiva celor care generează și gestionează marile baze de date contemporane, dimensiunea este, într-adevăr, un factor al cărui efect nu poate fi neglijat: dacă bazele tradiționale erau construite pentru interogări și răspunsuri punctuale, rapide, bazele masive de astăzi sunt folosite cu precădere pentru a se descoperi pattern-uri neanticipate și insesizabile cu mijloace intuitive sau, altfel spus, pentru a se genera *cunoaștere (knowledge discovery)* în absența vreunor interogații prealabile sau a unor supoziții transparente. De altfel, nou apăruta „știință a datelor” (*data science*) a și fost definită ca „studiul extragerii generalizate a cunoașterii din date”, iar întrebarea după care se ghidează cei care examinează baze de mari dimensiuni este cât se poate de simplă și de explicită: „Ce pattern-uri satisfac aceste date?” (cu observația că pattern-urile respective trebuie să aibă capacitate predictivă)¹³. Pattern-urile ar emerge, așadar, înainte de-a fi sesizată vreo rațiune pentru apariția lor, iar dacă se întâmplă să fie cumva și mai adecvate pentru formularea unor predicții corecte, de ce ne-am mai obosi să căutăm modele cauzale prin care să explicăm relațiile găsite?¹⁴

Adevărul este că, prin multitudinea de variabile și prin numărul imens de înregistrări, bazele de date actuale fac practic imposibilă formularea intuitivă a unor interogări despre a căror corectitudine să ne putem pronunța aprioric, și în acest context algoritmi precum cei pregătiți pentru învățarea automată pot identifica, *în locul nostru*, pattern-urile interesante și utilizabile în predicții¹⁵. O cercetare științifică în care prioritatea s-o dețină datele însele, și nu supozițiile noastre despre ele (de fapt, supozițiile despre situația care le prilejuiește și pe care ar „documenta-o”), ar presupune, așadar, să se acționeze în conformitate cu un principiu care ar suna

¹³ V. Dhar, *Data Science and Prediction*, Communications of the ACM 56, 12 (2013), pp. 64–66. Un specialist în analiza datelor urmărește să identifice „interesting and robust patterns that satisfy the data, where «interesting» is usually something unexpected and actionable and «robust» is a pattern expected to occur in the future” (p. 66). Ar fi util să amintim aici că informaticienii (sau „datiștii”) operează, de regulă, cu o ierarhie numită „piramida cunoașterii” (*DIKW*), la baza căreia ar sta datele (elemente abstracte), urmate de informații (elemente concatenate), cunoaștere (informații organizate) și înțelegere (*understanding, wisdom* – cunoaștere „interiorizată” și aplicată) – a se vedea, de exemplu: R.L. Ackoff, *From Data to Wisdom*, „Journal of Applied Data Analysis” 16, 1989: 3–9; R. Kitchin, *The Data Revolution. Big Data, Open Data, Data Infrastructures and Their Consequences*, London, Sage, 2014, Ch. 1.

¹⁴ „Why bother developing a detailed causal model if it is poor at prediction and, more important, likely to get worse over time due to «concept drift»?” (V. Dhar, *op. cit.*, p. 70). Asemenea situații și raționamente sunt cel mai des întâlnite în domenii care se ocupă de analiza comportamentelor umane, precum finanțe, marketing sau sănătate. În aceste domenii, identificarea algoritmică a unor pattern-uri în seturi complexe de date este, de regulă, o procedură mai simplă și mai relevantă decât modelarea teoretică a unui comportament (S. Succi, P. V. Coveney, *op. cit.*, p. 2).

¹⁵ V. Dhar, *op. cit.*, p. 68.

cam așa: „Acestea fiind datele, haideți să căutăm acum și o ipoteză”¹⁶. Dar asta ar însemna, în cele din urmă, acceptarea unor situații în care cercetarea științifică s-ar putea derula și fără a porni, ca de obicei, de la supoziții anticipate despre natura relațiilor dintre variabilele măsurate. Dificultatea acceptării unor asemenea situații vine chiar din faptul că nu mai pot fi subsumate manierei tradiționale de a concepe „normalitatea” în materie de cercetare științifică, în sensul că „normal” înseamnă să pornești de la o idee prealabilă (ipoteză, teorie) către niște date relevante, nicidecum invers¹⁷. Multitudinea datelor colectate în zilele noastre permite, de multe ori, testarea anumitor ipoteze formulate de cercetători, dar lucrul cu adevărat nou este acela că, prin intermediul unor algoritmi rulați pe calculatoare, devine totodată posibilă identificarea unor relații ce nu fuseseră avute în vedere în momentul când se demarase operațiunea de colectare a datelor respective. În asemenea cazuri, inteligența artificială funcționează ca o veritabilă „mașină de pus întrebări” fără vreo intervenție umană efectivă, iar prin generarea de ipoteze sau teorii ea se afirmă ca o creatoare de cunoaștere independentă¹⁸.

Poate că problema pe care o aduce în atenție „diluviul datelor” este generată, într-adevăr, de neclarități strecurate în chiar conceptul științei și de prejudecăți incluse de la început în ideea modernă de metodă științifică. Privind înapoi în istoria științei, unii observatori ai problemelor pe care le-a scos la lumină fenomenul Big Data au crezut că pot găsi soluția într-un concept de știință anterior celui modern, în „știința baconiană”. Inductivismul baconian nu este însă evaluat unitar de către istoricii și filosofii științei. Deși multă vreme Francis Bacon a fost privit drept „părintele științei moderne”, istorici precum celebrul Alexandre Koyré sau filosofi de talia lui Karl Popper au respins cu hotărâre o asemenea interpretare¹⁹. Așa cum

¹⁶ V., spre exemplu, articolul lui D. B. Kell și S. Oliver, *Here Is the Evidence, Now What Is the Hypothesis? The Complementary Roles of Inductive and Hypothesis-driven Science in the Post-genomic Era*, „BioEssays” 26, 2003: 99–105.

¹⁷ Învățarea automată, remarcă Dhar (*op. cit.*, p. 67), presupune o răsturnare a imaginii noastre tradiționale despre felul cum ar trebui să se desfășoare o cercetare științifică: „Given a large trove of data, the computer taunts us by saying, «If only you knew what question to ask me, I would give you some very interesting answers based on the data.» Such a capability is powerful since we often do not know what question to ask”.

¹⁸ „By initiating interesting questions and refining them without active human intervention, it becomes capable of creating new knowledge and making discoveries on its own” (R. Agarwal, V. Dhar, *Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research*, „Information Systems Research” 25 (3), 2014, p. 444).

¹⁹ Alexandre Koyré (*Études galiléennes*, Paris, Hermann, 1966, p. 12) era convins că rolul lui Bacon în istoria revoluției științifice „a fost perfect neglijabil”: „«Bacon initiateur de la science moderne» est une plaisanterie [...]. En fait, Bacon n’a jamais rien compris à la science. Il est [...] crédule et totalement dénué d’esprit critique. Sa mentalité est plus proche de l’alchimie, [...] de celle d’un primitif ou d’un homme de la Renaissance que de celle d’un Galilée”. Sunt însă câteva decenii de când această opinie este contestată, chiar și atunci când se analizează o scriere atât de ciudată – privită prin ochii cititorului de știință al zilelor noastre – precum *Sylva Sylvarum*. A se vedea, de exemplu, P.R. Anstey, *Philosophy of Experiment in Early Modern England: The Case of Bacon, Boyle and Hooke*, „Early Science and Medicine” 19 (2), 2014: 103–132; D. Jalobeanu, *Disciplining Experience: Francis Bacon’s Experimental Series and the Art of Experimenting*, „Perspectives on Science” 24, 3, 2016: 324–342.

se știe, pentru a ne feri de „idoli” Bacon propunea să pornim pe „calea adevărată, dar încă neîncercată”, cea care, în opoziție cu deductivismul clasic (inclusiv cu ceea ce avea să se numească apoi metoda ipotetico-deductivă) „scoate propozițiile din datele simțurilor și din faptele particulare, ridicându-se continuu și gradual, pentru a ajunge, în cele din urmă, la propozițiile cele mai generale”²⁰. Termenul „știință baconiană” a fost însă parțial deturnat de comentariile lui Thomas Kuhn, care a folosit această expresie pentru a diferenția științele fizice zise „clasice” (astronomia, statica și optica) de „studiile baconiene” slab conceptualizate și structurate, interesate de investigarea unor fenomene fizice precum căldura și magnetismul (sau a unor fenomene chimice, în fazele incipiente ale chimiei)²¹. O știință fără ipoteze ar fi, oare, și o știință ferită de prejudecăți, fără acei „idoli” de care dorea să ne elibereze Francis Bacon și de care ar fi vrut să se păzească până și Newton, cum o confirmă celebra lui remarcă *Hypotheses non fingo?*

În orice caz, situația favorizată de apariția Big Data este de multe ori asemănată cu cea care a inspirat inductivismul baconian, privit ca o formă de empirism extrem, din perspectiva căruia cuvântul „ipoteză” era întotdeauna sinonimul prezumțiilor hazardate și fanteziste. De pe poziția unui astfel de empirism se consideră că marile baze de date ar oferi o cuprindere relativ completă a unui domeniu, ar elimina necesitatea de-a se formula ipoteze, modele sau teorii și ar permite acea analiză statistică „agnostică” cu ajutorul căreia datele ar putea să „vorbească” prin ele însele. Desigur, sub aspect practic un asemenea empirism dus la extrem, prin maxima prioritizare a datelor, ar fi o imposibilitate, de vreme ce însăși alegerea datelor ce se vor culege presupune situarea într-un context teoretic sau măcar cultural, context ce reapare inevitabil și în generarea pattern-urilor, iar datele nu pot „vorbi” decât într-un cadru uman²². În plus, corelațiile între variabilele cuprinse în bazele de date sunt foarte problematice, în sensul că cele mai multe dintre ele sunt înșelătoare (*spurious correlations*). Din acest motiv se poate spune că informația în cantități prea mari tinde să se comporte la fel de nesatisfăcător ca și informația în cantități prea mici, și de aceea „metoda științifică poate fi

²⁰ F. Bacon, *Noul Organon*, trad. de N. Petrescu și M. Florian, București, Editura Academiei RSR, 1957, p. 38 (I.XIX).

²¹ Thomas Kuhn (*Tensiunea esențială. Studii despre tradiție și schimbare în știință*, trad. de A. Florea, București, Editura Științifică și Enciclopedică, 1982) crede că, deși „baconismul a contribuit puțin la dezvoltarea științelor clasice, el a dat naștere unui mare număr de noi domenii științifice” (p. 96). Fiind însă incapabile să producă „predicții amănunțite”, științele baconiene au rămas subdezvoltate în secolul al XVII-lea și în mare parte a secolului al XVIII-lea, ceea ce amintește de situația pe care o întâlnim astăzi în științele sociale (pp. 96–97). Așadar, deși „inițial a fost sărăcăcios în ceea ce privește rezultatele conceptuale, programul baconian a inaugurat totuși unele dintre științele moderne mai importante” (p. 165).

²² „Making sense of data is always framed – data are examined through a particular lens that influences how they are interpreted. Even if the process is automated, the algorithms used to process the data are imbued with particular values and contextualized within a particular scientific approach” (R. Kitchin, *Big Data, New Epistemologies and Paradigm Shifts*, „Big Data & Society”, April–June 2014: 1–12. DOI:10.1177/20539517145284812014 – p. 5).

îmbogăţită prin exploatarea minieră a bazelor de date imense, dar nu poate fi înlocuită de aceasta”²³.

Inductivismul baconian mai este însă important și prin faptul că sugerează posibilitatea algoritmizării și automatizării descoperirii științifice. Pentru că nu putem avea încredere în „mersul nativ și de la sine al minții”, spune Bacon, trebuie să acceptăm că „rămâne numai o singură salvare și însănătoșire, anume ca întreaga operă a intelectului să fie refăcută, iar mintea însăși, din capul locului, [...] să fie îndrumată la fiecare pas; și lucrul să fie îndeplinit *ca la mașină*”²⁴. Se pare că abia astăzi a devenit realmente posibilă punerea în practică a sugestiei baconiene privind „automatizarea” descoperirii științifice. Într-adevăr, cel puțin în domeniul studiilor biologice s-au făcut progrese remarcabile în crearea unor „roboți creatori de știință” (*robot scientists*). Un asemenea robot este un sistem fizic de automatizare a unor lucrări de laborator, sistem care, bazându-se pe inteligența artificială, execută cicluri de experimentare științifică. Faptul remarcabil este acela că un asemenea robot *generează în mod automat ipoteze* pentru a explica anumite observații, concepe experimente pentru testarea ipotezelor, le testează efectiv și interpretează rezultatele obținute, iar la sfârșit reia ciclul²⁵.

În opinia lui Rob Kitchin, această formă de empirism cu origini baconiene ar fi potrivită în domenii mai puțin structurate sub aspect conceptual și teoretic, adică în acele domenii (sănătate, științe sociale, științele pământului) în care nu dispunem încă de „teorii solide”, dar unde avem, în schimb, cantități uriașe de date. Inductivismul ar fi însă mai puțin folositor în domeniul științelor bine structurate, adică acolo unde se dispune de teorii „complete” care ne permit explicarea integrală a unui fenomen – domeniul științelor „tari”, unde predicția este sinonimă cu explicația²⁶.

Există însă și o formă mai atenuată de manifestare a primatului datelor, concretizată în metoda cunoscută sub numele de *cercetare dirijată de date* (*data-driven research*). Așa cum propune o analiză făcută de Rob Kitchin, se poate spune că prin acest termen este desemnată o formă de cercetare mai apropiată de ceea ce recunoaștem încă a fi metoda științifică modernă, diferența venind din folosirea unei „combinații hibride de abordări abductive, inductive și deductive” pentru

²³ C. Calude, G. Longo, *The Deluge of Spurious Correlations in Big Data*, „Foundations of Science” 2016. DOI 10.1007/s10699-016-9489-4

²⁴ F. Bacon, *op. cit.*, p. 30 – subl. mea G.N. Algoritmul gândit de el ar avea trei componente esențiale: „First, observations have to be collected and integrated into the total corpus of knowledge. Second, the new observations are used to generate new hypotheses. Third, the hypotheses are tested through carefully designed experiments” (A. Alkhatieb, *Can Scientific Discovery Be Automated?*, „The Atlantic” 25 Apr. 2017, theatlantic.com/science/archive/2017/04/can-scientific-discovery-be-automated/524136).

²⁵ R.D. King et al., *The Automation of Science*, „Science” 324, 3 APRIL 2009, pp. 85–89.

²⁶ „In the «hard» sciences, where models can be assumed [...] to be complete, there exists the possibility of extracting causal models from large amounts of data. In other fields, large amounts of data can result in accurate predictive models, even though no causal insights are immediately apparent” (V. Dhar, *op. cit.*, p. 71).

înțelegerea unui fenomen. Deosebirea majoră față de abordarea ipotetic-deductivă se manifestă prin aceea că ipotezele și intuițiile cu valoare integratoare sunt generate „din date” mai degrabă decât „din teorie”. Se încearcă astfel o mai atentă încorporare a proceselor inductive în cercetare, înainte de folosirea unor procedee deductive, dar fără a se urmări formularea unor „explicații” prin simplă inducție. Strategia epistemologică generală implicată aici ar consta deci în folosirea unor „tehnici dirijate de descoperire a cunoașterii pentru a se identifica potențiale întrebări (ipoteze) care să merite examinarea și testarea ulterioară”²⁷. În aceste condiții s-ar putea afirma că știința dirijată de date poate fi înțeleasă ca o „versiune reconfigurată a metodei științifice tradiționale care oferă o nouă cale de construire a teoriei”²⁸. Prin potențialul ei de-a extrage sau genera intuiții suplimentare față de cercetările orientate de teorie (sau de cunoaștere, cum se mai spune), sunt șanse ca această cercetare dirijată de date să se impună ca noua paradigmă a metodei științifice – o a patra paradigmă a științei.

3. Fenomenul „Big Data” este privit, uneori, doar ca una dintre manifestările „revoluției digitale” generată de progresul științei și tehnologiei informației, alături de el fiind menționate evoluțiile din inteligența artificială, învățarea automată (*machine learning*), simularea sau calculul în rețele (*cloud computing*)²⁹. Din punctul de vedere care ne interesează aici, trăsătura lor comună ar fi tentația de a se depăși, pe cale tehnologică, standardele uzuale de raționare științifică prin promovarea unui „stil de raționare post-uman” (*i.e.*, artificial) sau „străin” (*alien reasoning*), direcționat tocmai spre transgresarea tehnologică a capacității inferențiale specifică omului³⁰. În fine, ar mai trebui adăugată aici și ideea automatizării descoperirii științifice prin realizarea acelor „roboți creatori de știință”, care trimit cu gândul la tema „singularității tehnologice”, a acelei evoluții tehnologice exponențiale despre care vorbește scriitorul SF Vernor Vinge.

Așa cum remarcă Thomas Nickles, se pare că revoluția digitală contemporană a reușit să evidențieze anumite limitări și disfuncții care subminează adaptabilitatea concepțiilor noastre despre cunoaștere și cercetarea științifică la provocările noilor tehnologii. Am văzut mai sus că, pornind de la problemele pe care le generează fenomenul Big Data, ajungem să repunem în discuție, rând pe rând, concepte și teme

²⁷ Kitchin, *Big Data...*, pp. 5–6. În concluzie: „whilst it is clear that Big Data is a disruptive innovation, presenting the possibility of a new approach to science, the form of this approach is not set, with two potential paths proposed that have divergent epistemologies – empiricism, wherein the data can speak for themselves free of theory, and data-driven science that radically modifies the existing scientific method by blending aspects of abduction, induction and deduction” (p. 10).

²⁸ Kitchin, *Big Data...*, p. 6.

²⁹ T. Nickles, *Alien Reasoning: Is a Major Change in Scientific Research Underway?*, „Topoi” 2018, doi.org/10.1007/s11245-018-9557-1

³⁰ Deși au fost inspirate de rezultate și supoziții privind modul cum funcționează creierul și gândirea umană, tehnicile folosite în inteligența artificială și învățarea automată „are designed precisely not to limit computation to human-style, serial, explicit reasoning. The traditional human standard is too restrictive, too limiting in inferential power” (*ibidem*).

epistemologice tradiţionale precum „contextul descoperirii” şi „contextul justificării”, corelaţia şi cauzalitatea, explicaţia, unitatea ştiinţei, monismul metodologic, ba însuşi conceptul nostru de ştiinţă şi, în cele din urmă, ideea metodei ştiinţifice. Într-o epocă a pluralismului privit ca un fel de relativism „virtuos”, ideea revigorării unui anume tip de „ştiinţă baconiană”, a unui inductivism reclamat pur şi simplu de evoluţiile tehnologice care au loc în zilele noastre, nu mai poate trece drept o exagerare. Discuţiile pe marginea efectelor pe care le va avea „diluviul datelor” asupra concepţiei noastre epistemologice mi se pare că se încadrează perfect în acea tot mai răspândită viziune pluralistă asupra ştiinţei şi asupra manierelor filosofice de a ne raporta la ea.