Invention vs. Discovery
A critical discussion

Carlotta Piscopo and Mauro Birattari

IRIDIA, Université Libre de Bruxelles, Brussels Belgium
piscopo@iridia.ulb.ac.be, mbiro@ulb.ac.be

Abstract. The paper proposes an epistemological analysis of the dichotomy discovery/invention. In particular, we argue in favor of the idea that science does not discover facts by induction but rather invents theories that are then checked against experience.

1 Introduction

Reasoning about the status of knowledge has always been integral to both science and philosophy: What is the path that lead from experience to theories?

The first modern answer to this question was given in the early age of modern science: According to Francis Bacon, knowledge is obtained directly from experience. More precisely, Nature is ruled by laws and the task of the scientist is simply to discover and describe them. For Bacon, science is an inductive process. Such position, henceforth termed *discoverist*, entails the idea that scientific knowledge grows linearly and cumulatively.[1] Starting from the 19th century a new trend in science deeply undermined this idea. The introduction of non-Euclidean geometries first, and then of relativity theory and quantum mechanics dared question theories that had been considered sure for centuries. In particular, the fact that mechanics—queen of the sciences—had to undergo a radical revision created a major shock in all scientific domains. Such shock clearly touched also the domain of philosophy of science where serious doubts were raised about the cumulative view of science and about the very idea of scientific discovery.

In particular, the concerns about induction previously raised by David Hume [5] came to a new life and were followed in the early 20th century by a large number of epistemological analysis that re-proposed similar issues in a more modern language. Among the others, Pierre Duhem [2] in a critical analysis of Newton's mechanics questioned its inductive nature. Further, Bertrand Russell [14] with the famous argument of the *inductivist turkey* raised doubts about the reliability of a theory based on induction. Finally, Karl Popper [12, 13] rejected firmly any inductivist basis for science and proposed, for the first time in a mature form, the alternative view that considers theories as conjectures.[2]

[1] The view of science as a *cumulative process* was described by Thomas Kuhn [6] as opposed to his view in which science is a process composed by irreconcilable steps.

[2] In *Logic der Forschung* [12] Popper maintains that scientists *invent*—rather than *discover*—laws and then they check them against experience. In this respect, the title

We will call this second epistemological position *inventionist* and the rest of the paper will focus on the dichotomy *inventionism/discoverism*.

Notwithstanding the strong and authoritative criticisms raised against induction, the inductivist—and therefore discoverist—positions boldly reemerged in artificial intelligence. Starting from the first works on expert systems, the very idea was put forward that it is possible to build programs that make scientific discoveries by induction from data. The expert system BACON.1 is a milestone in machine learning and it is an important example of how the inductivist and discoverist idea has been implemented [8]. Further, the discoverist position emerged through several editorial events such as a special issue of *Artificial Intelligence* [15] and our conference *Discovery Science*. Altogether, they prove the interest of the AI community for the concept of discovery and for the related inductivist view of science. It is worth noticing, however, that recently some sectors of the machine learning community seem to have definitely switched to an inventionist position.[3]

Our personal background and our specific research interests lead us to accept an inventionist position. Within the conference DS-2002 we intend to take the role of the devil's advocate bringing to the fore in this community the idea that science is about invention! Clearly we are not animated in this discussion by any sake of argument. On the contrary, with this extended abstract and with our presence at the poster session of DS-2002, we wish to start a fruitful and open discussion. In our opinion this discussion could help refining our relative position in full awareness of their philosophical and epistemological implications.

2 Discovery in science

The discoverist idea is based on the assumption that observation alone is enough to find the laws of Nature. According to this view, an accurate collection and organization of data lets immediately emerge the intrinsic regularities of Nature.

The idea that laws are already in Nature and that science is about discovering them, traces back to the ancient and medieval philosophy. According to such idea, theories ultimately reflect the very structure of reality. Francis Bacon,

The logic of scientific discovery of the English translation is particularly misleading and seems to suggest the opposite idea: indeed the German *forschung* means literally *research* rather than *discovery*.

[3] All non-parametric statistical methods such as bootstrap [3] and cross-validation [16] do not rest on the hypothesis that the real system under observation belongs to the model space. Indeed, if the system does not belongs to the model space the learned model cannot coincide with the system itself and therefore no *discovery* is possible. In such a case, the learned method can be at best an approximation of the system and remains something *ontologically* distinct from the latter. The learned model can be therefore considered only as a useful *invention*. Vladimir Vapnik [18] is even more explicit about the dichotomy invention/discovery: he cites directly Popper and the concept of falsification. In some sense, the *VC-dimension*, key concept of Vapnik theory, can be seen as a modern and mathematically rigorous version of the concept of dimension of a theory discussed by Popper [12].

father of the experimental method, embodies such an idea. Bacon's picture of science rests upon the regulative idea that natural laws have to be extracted only from pure empirical data. Accordingly, he conceives in his *Novum Organum* [1] two distinct phases that should characterize the experimental method. First the experimenter must put aside all theoretical anticipations that Bacon [1] colorfully calls *idola*.[4] Second, in the proper experimental phase, data are collected and organized in what he calls *tabulae*—the forerunners of modern *databases*.

The assumption that natural laws can be extracted simply from experimental data raised the most controversial issue in the whole history of epistemology: the problem of induction. Hume [5] is the first thinker that openly and strongly maintains that empirical laws are not logically entailed by observed data, but are rather subjective conjectures originating from the habit to see regularities in repeated events. A century after Hume, John Stuart Mill [9] argues again in favor of the inductivist idea but, aware of Hume's argument, he justifies induction through the extra-scientific assumption that Nature is ordered by deterministic laws.

The discoverist idea defines the role of science while preserving its *objectivity*: science is about diving into empirical data for finding the laws of Nature. The price to pay is that the discoverist position has to deal with the problem of induction. The determinism of Nature postulated by Mill [9] is an attempt to solve the problem. Yet, it raises other concerns for its clear metaphysical flavor. As Whitehead [19] pointed out, the belief in the deterministic order of nature is nothing but the reinterpretation of the medieval belief in a rational God.[5]

3 Invention in science

As pointed out in Section 1, the discoverist idea emerged as a result of great historical changes. Non-Euclidean geometries together with relativity theory questioned basic concepts such those of space and time that according to Newton and Kant enjoyed the property of being absolute and necessary. Besides that, relativity theory and quantum mechanics showed that the Newtonian mechanics, that had been considered as the true description of the universe, was just an approximation leading in some circumstances to incorrect predictions. The scientific

[4] The term *idolum* comes from the Greek *eidolon* which means image or phantom. By adopting this term, Bacon makes clear that in his views theoretical ideas are misleading and they prevent from reaching pure empirical knowledge that alone leads to the discovery of the laws of Nature.

[5] This idea is not so inconsiderate as it seems since metaphysical ideas are behind many scientific disciplines, classical mechanics included. Newtonian mechanics rests upon the idea of an "intelligent and powerful Being" that is ultimately responsible of the order of Nature [10]. On the other hand, Leibnizian mechanics supposes that the world we experience is nothing but the one that God chooses as the best among many possible others. Through the *principle of least action*, such idea carries on to the Euler-Lagrange theory, to the Hamilton-Jacoby theory and ultimately to all contemporary formulations of mechanics [7].

and epistemological crisis opened by the refutation of Newton's mechanics undermined the very idea that scientific theories are discovered and justified, once and forever, by inductive processes. As opposed to the discoverist epistemology, a different view was proposed in the early 20th century, according to which scientific theories are bold speculations put forward and maintained as "true" until they resist to the test of experience. Accepting a theory as true *unless proved false* might seems rather weak but, as we will see presently, this appeared as the only way for skipping the problem of induction as raised by Hume.

Embodying the inventionist epistemology, Popper puts forward the idea that scientific theories neither come out directly from experience nor are definitively verified by it. Clearly Popper accepts that scientific theories can be *suggested* by observation but he firmly denies that experience alone can logically justify a theory. Popper [13] provocatively defines induction as a "myth". According to Popper, when we observe we have an interest, a problem to solve, a viewpoint, and a theory for interpreting the world, which make us selectively search in the huge amount of data obtained from observation [13]. Popper clarifies that: *scientific theories were not the digest of observation, but that they were inventions–conjectures boldly put forward for trial, to be eliminated if they clashed with observation*—[13]. Such conjectures, that are neither obtained by induction nor are verified definitively, can however be falsified by experience.

It is on the logical aspect of the theory of knowledge that Popper focuses his attention: *The theory to be developed* [...] might be described as the theory of *the deductive method of testing* [...] I must first make clear the distinction between the *psychology of knowledge* which deals with empirical facts, and the *logic of knowledge* which is concerned only with logical relations. [...] the work of the scientist consists in putting forward and testing theories. The initial stage, the act of conceiving or inventing a theory, seems to me neither to call for logical analysis nor to be susceptible of it. The question how it happens that a new idea occurs to a man [...] may be of great interest to empirical psychology; but it is irrelevant to the logical analysis of scientific knowledge. [...] Accordingly I shall distinguish sharply between the process of conceiving a new idea, and the methods and results of examining it logically.–[12]. Popper then focuses his discussion on the *logical examination of a theory* and therefore on falsification and refutation of conjectures rather than on the origin of the latter: It has been sharply noted [4] that, in spite of the title, in Popper's *Conjectures and Refutations* [13] the term 'conjecture' does not even appear in the index!

By introducing what he calls the *asymmetry* between verifiability and falsifiability, Popper explains that while a scientific theory cannot be derived from observations, it can be contradicted by them [12]. This is done by a deductive procedure called in logic *modus tollens*, through which we can argue *from the truth of singular statements to the falsity of universal statements*—[12]. Eliminating the recourse to induction as a mean to explain both how theories are obtained and how they are tested, Popper solves the problem of induction.

Other scientists participated in the debate on induction. The great physicist and mathematician Henri Poincaré is particularly representative. He elaborated an epistemological conception of science, called *conventionalism*: no scientific theories can aspire to obtain the status of true representation of the world. They are simply useful *conventions* that science puts forward and uses only because they yield good predictions [11]. Scientific theories are therefore simply stipulations that the scientific community decides by agreement to assume, or eventually abandon, according to their utility.[6]

The problem of induction skipped, another problem emerges: if theories are seen as inventions [12, 13], or alternatively as conventions [11], they loose their character of *objectivity*. However, Popper's and Poincaré's viewpoints do not entail that theories are completely arbitrary. On the contrary, according to Popper and Poincaré theories are inter-subjective: the objectivity of scientific theories comes from the possibility of being "inter-subjectively tested" by scientists [12]. The decision about the destiny of theories is left to the scientific community that is in charge of testing the predictions these theories allow [11].

4 Conclusion

Assuming that science is about *discovering* exposes to the problem of induction. However, if experience is assumed to be the only basis on which scientific theories rest, the objectivity of science can be maintained.

On the other hand, assuming the opposite view according to which science is about *inventing* theories protects from the problems related to induction since the source and the justification of theories do not rest ultimately upon experience, but upon a decision. According to this view, theories are conjectures. In this sense, the objectivity of science cannot be maintained anymore. Nevertheless, scientific theories are not arbitrary since they must be inter-subjectively testable, and possibly falsified.

Acknowledgments: Mauro Birattari acknowledges support from the Metaheuristics Network, a Research and Training Network funded by the Commission of the European Communities under the Improving Human Potential programme, contract number HPRN-CT-1999-00106. The information provided is the sole responsibility of the authors and does not reflect the Community's opinion. The Community is not responsible for any use that might be made of data appearing in this publication.

References

[1] F. Bacon. *Novum Organum.* 1610. Available in: *The works*, B. Montague (ed.), Parry & MacMillan, Philadelphia, PA, USA, 1854.

[2] Pierre Duhem. *La Théorie Physique: son Object et sa Structure.* 1906. Available as: *The Aim and Structure of Physical Theory*, Princeton University Press, Princeton, NJ, USA, 1991.

[6] Beside the predicting power, other criteria regulate the choice of a theory. Such further criteria, which are typically extra-evidential, might be for instance simplicity (Occam's razor) or conservatism.

[3] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap.* Chapman and Hall, New York, NY, USA, 1993.

[4] P. A. Flach. *Conjectures: an inquiry concerning the logic of induction.* PhD thesis, Katholieke Universiteit Brabant, Tilburg, The Netherlands, 1995.

[5] D. Hume. *A Treatise of Human Nature.* 1739. Available as: E. C. Mossner (ed.), Penguin Books, London, United Kingdom, 1986.

[6] T. Kunh. *The Structure of Scientific Revolutions.* The University of Chicago Press, Chicago, IL, USA, 1996: 3rd edition, 1962.

[7] C. Lanczos. *The Variational Principles of Mechanics.* Dover Publications, New York, NY, USA, 1986.

[8] P. Langley, H. A. Simon, G. L. Bradshaw, and J. M. Zytkov. *Scientific discovery. Computational Explorations of the Creative Processes.* MIT Press, Cambridge, MA, USA, 1987.

[9] J.S. Mill. *A System of Logic.* Longmans Green, London, United Kingdom, 1843.

[10] I. Newton. *Mathematical Principles.* 1713. Available as: F. Cajori (ed.), University of California Press, Berkeley, CA, USA. 1946.

[11] H. Poincaré. *La Science et L'Hypothèse.* 1903. Available as: *Science and Hypothesis*, Dover Publications, New York, NY, USA, 1967.

[12] K. Popper. *Logik der Forschung.* 1935. Available as: *The Logic of Scientific Discovery*, Routledge, London, United Kingdom, 1999.

[13] K. Popper. *Conjectures and Refutations.* Routledge and Kegan Paul, London, United Kingdom, 1963.

[14] B. Russell. *The Problems of Philosophy.* Williams and Nogate, London, United Kingdom, 1957.

[15] H. A. Simon, R. E. Valdés-Pérez, and D. H. Sleeman. Scientific discovery and simplicity of method. *Artifical Intelligence*, 91(2):177–181, 1997. Editorial of a Special Issue on Scientific Discovery.

[16] M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society B*, 36(1):111–147, 1974.

[17] R. S. Sutton and A. G. Barto. *Reinforcement Learning. An Introduction.* MIT Press, Cambridge, MA, USA, 1998.

[18] V. N. Vapnik. *Statistical Learning Theory.* John Wiley & Sons, New York, NY, USA, 1998.

[19] A. N. Whitehead. *Science and the Modern World.* Cambridge University Press, Cambridge, United Kingdom, 1926.