THEORETICAL PREDUCTION AND COMPUTATIONAL SCIENTIFIC DISCOVERY

ANDRÉS RIVADULLA[1]

ABSTRACT: Can deductive reasoning be applied to the context of discovery of mathematical physics? I term *preduction* the method of reasoning that, starting from the available theoretical background as a whole, allows for the anticipation of previously unknown results, provided that the combination of already accepted results of different theories – those taken as premises of the deductive reasoning – is compatible with dimensional analysis. This is the method by which many hypotheses, laws, and theoretical models are introduced into physics. *Computational preduction* further extends the possibilities of machine learning, which in the past forty years has informed computational systems, implementing both the rediscovery of empirical laws and the automated discovery of equations in data bases. Automated preduction would also facilitate scientific creativity in theoretical physics.
KEY WORDS: scientific discovery, theoretical preduction, computational discovery, computational preduction

1. Problem solving and induction

Two years after the publication of Popper's *Objective Knowledge*, a book in which he develops his anti-inductivist stance, which states that we can never rationally justify our belief in the truth of a theory, or in its being probably true, Herbert Simon (1974, 330) proclaimed the close relationship between induction and problem solving: "from a logical standpoint the processes involved in problem solving are inductive, not deductive." More specifically: "The problem solving process is an information gathering process as much as it is a search process. ... The process for using this information to steer the search are generally processes of inductive inference. Being inductive, they do not provide certainty, but have only heuristic value in guiding the search and making it efficient."

Since both positions are not compatible with each other, one possible outcome is to assume that the logical illegitimacy of induction is not decisive for the methodology of science. Langley *et al.* (1987, 16-17.My italics, A. R.) are perfectly aware of the weakness of inductive inferences. In an argument reminiscent of Popper, they maintain that:

> "It is well known that no universally quantified generalization can be verified decisively by a finite set of observations. Not only may subsequent observations overthrow the verification, but it will never be the case that the generalization that

1 Projects FFI2009-10249 and AIB2010PT-00106 supported by the Spanish Science and Innovation Ministry.

> has been found to fit the observations is unique. Other generalizations, known or not, will describe or explain the data just as well as the one being considered."

But, instead of concluding, with Popper, that induction simply does not exist, they conclude that *science solves these problems of certainty and uniqueness by ignoring them*. In relation to the *flawed assumption* that we are seeking an *infallible* method of scientific discovery, they affirm:

> "The discovery process solves the problem of infallibility, as it does the problems of certainty and uniqueness of generalizations, by ignoring it. ... We require neither a guarantee that solutions, in the form of laws, can be found, nor a guarantee that the laws can be strongly verified." (*op. cit.*, 303)

Insisting on this idea, Langley *et al.* (*op. cit.*, 16-17.My italics, A. R.) maintain that

> "The problem of discovery is precisely to find some generalization that is parsimonious and consistent with all the given data (within some criterion of accuracy). At the point of discovery, no claim is made that the generalization will survive the test of additional evidence, or that another generalization will not be found that is equally adequate for describing the original data. (...) If even one generalization or explanation can be found to fit a body of data reasonable well, it will usually be accepted, at least until a competitive law or explanation comes along or until new important data are found that do not fit at all. When two or more explanations for the same data or for overlapping data are discovered, there arise the new tasks of evaluating them relative to one another and (usually) discovering new laws that reconcile the discrepancies. In all these circumstances, the first and vital task is to discover laws. While that is going on, the existence of competitors, actual or potential, is simply irrelevant. *The concern that the philosophy of science has had for the certainty and uniqueness of inductions stems from a misconception of the nature of science.* ... Laws, or supposed laws, will often be discovered (a happy moment). They will often be refuted by later evidence (a sad moment), and they will sometimes have to contend with rival laws that explain the same data (a perplexing moment)."

Thus Langley's *et al.*, (*op. cit., ibid.*) position on the Humean *induction problem* is:

> "The function of verification procedures is not to provide scientists with unattainable certainty or uniqueness for discoveries, but to inform them about the risks they are running in committing themselves to hypotheses that have been discovered and to provide guidance that may enhance their chances of making relatively durable discoveries."

Popper's issues with the induction problem do not seem to be justified.

Moreover Langley *et al.* (*op. cit.*, 39 and 40, respectively) are very far from endorsing a naive verificationism:

> "The notion that theories are to be validated by deriving them deductively from observations – usually called *naive verificationism* – is clearly invalid." "Of course, many scientific papers contain statements about evidence 'supporting', 'strengthening', or 'weakening' a theory. ... But such statements fall far short of constituting quantitative claims of 'degree of justifiable belief'. There does not seem to be any essential requirement in science for such claims."

As to the relationship between the context of justification and the context of discovery, which for them are not separate and distinct questions (see Rivadulla 2012), Langley *et al.* (*op. cit.*, 58-59) conclude:

> "First, a new discovery ... is generally developed through a step-by-step search guided at each step by data. Second, at each step in this process, the developing hypothesis is guaranteed to be confirmed by the data examined thus far. Third, since only a finite body of data is in question, the verification can be complete (that is, compatibility with the data can be definitely confirmed), although uniqueness cannot be guaranteed. Fourth, because of the incremental character of the process, the discovery, once arrived at, has already achieved a considerable degree of confirmation, although it still may fail when tested against new data."

Paul Thagard (1988, 27-28) also points to the limitations of induction:

> "inductive inference [is] inference to conclusions that do not follow necessarily from what is already known. Induction ... can lead to conclusions that are not only useless but false. At least in deduction, what follows from true premises has to be true, whereas induction unavoidably involves a kind of leap. You may have observed myriad instances of copper conducting electricity, but generalizing that all copper conducts electricity still introduces uncertainty."

Thus, according to Thagard, the use of induction cannot go beyond the production of rules for problem solving and explanation:

> "in making inductions, we do not want to be concerned only with reaching well-warranted conclusions: we want to produce rules that will be useful for future use in problem solving and explanation."

This position is in complete agreement with Simon's (1995, 175. My italics, A. R.):

> "Nor is exact truth the realistic goal of discovery. ... The task of discovery is to create theories that explain the data 'well enough', with the hope that tomorrow they can be explained better. ... We call a problem solution a discovery to the extent that it is both valuable (and/or interesting) and novel. So *an empirical theory of discovery is a theory of how novel and valuable (or interesting) problem solutions are in fact found*."

2. Automated induction and computational scientific discovery

To claim that scientific discovery proceeds computationally points to the existence of methods, systems or programs that allow for its automation. As Saso Dzeroski *et al.* (2007, 1) affirm: "Research on computational scientific discovery aims to develop computer systems which produce results that, if a human scientist did the same, we would refer to as discoveries."

The computational science of scientific discovery is grounded in the development of computational systems that from their very beginning have been able to reproduce, replicate, simulate, or rediscover important laws of the history of science. Such systems or programs, usually written in LISP or in PROLOG, did appear during the 1960s, and since the 70s they have become well established. From the perspective of the twenty-first century, referring to the set of programs collectively called BACON, Dzeroski *et al.* (2007, 8.My italics, A.R.) affirm that they implemented

> "data-driven induction of descriptive laws and were demonstrated on historical examples. *Together, they provided the first compelling evidence that computational scientific discovery was actually possible.*" "The early work on computational discovery focussed on reconstructions from the history of science that were consistent with widely accepted theories of human cognition. This was an appropriate strategy, in that these examples let researchers test their methods on relatively simple problems for which answers were known." (*op. Cit.*, 9)

Langley's *et al.* (1987, 20.My italics, A. R.) book is mainly concerned with "*the induction of descriptive and explanatory theories from data.*" In order to implement this task, these authors employ the computer program BACON, which is capable of inductively producing scientific discoveries from empirical data. Referring to BACON, Shrager and Langley (1990, 10) claim: "BACON ... focused on the induction of numerical laws from experimental data." And Thagard & Nowak (1990, 28) affirm that "the BACON program takes numerical values as inputs and generates equations that fit those values." Also, Herbert Simon (1992, 8.My italics, A. R.) insists on the pure inductive character of BACON:

> "*A series of computer programs, collectively named BACON, has successfully simulated the process whereby a substantial number of important laws of eighteenth and nineteenth century chemistry were induced from data.*"

For instance, BACON is able to obtain Kepler's Third Law inductively from data. But it can also obtain Joseph Black's laws of temperature equilibrium, the law of conservation of momentum, and many others. It does not only find laws, but is also able to introduce new theoretical concepts, like inertial mass, specific heat, atomic weight, etc. Simon (1995, 177) also recognizes BACON's capabilities as a data-driven automated discovery program:

> "It can be claimed that BACON's process resembles closely the process that human scientists have used in a number of important historical cases where little or no theory was available to guide the search for laws: hence 'data-driven discovery'."

3. Computational preduction

Computational scientists repeatedly assert that they have filled the gap left behind by philosophers of science, and they reproach them for abandoning the context of discovery (see Rivadulla 2012). Is the moment now ripe for the philosophers of science to regain the lost ground and to advance new possibilities in computational science? In order to answer this question, it is necessary to extend the spectrum of possibilities of scientific discovery. In this respect, both a wider concept of 'scientific discovery' and a closer attention to real scientific methods are unavoidable. In particular we must investigate the possibility of new strategies for the introduction of novel ideas into science, beyond standard ampliative inferences, such as induction and abduction.

As to the first issue, Dzeroski *et al.* (2007, 3-4.My italics, A. R.) define scientific discovery – briefly, and without further reference to exclusive methods of research – as

> "the process by which a scientist creates or finds some hitherto unknown knowledge, such as a class of objects, an empirical law, or an explanatory theory. ... A defining aspect of discovery is that the knowledge should be *new and previously unknown.*"

When addressing the issue of non-standard ways of introducing new ideas into science, philosophers of science have maintained different perspectives. Peter Medawar (1974, 289) pointed to the weakness of deductive reasoning in scientific methodology:

> "The weakness of the hypothetico-deductive system, insofar as it might profess to offer a complete account of the scientific process, lies in its disclaiming any power to explain how hypotheses come into being."

For his part, Thomas Nickles (2008, 446) guesses at the applicability of deductive reasoning in discovery tasks:

> "it is worth noting that even an ordinary deductive argument need not be sterile: it may be epistemologically ampliative even though it is not logically ampliative, for we are not logically omniscient beings who see all the logical consequences of a set of propositions."

And Elie Zahar (1983, 244-245.My italics, A. R.) has unambiguously advanced the thesis of the deductive character of scientific discovery:

> "*the process of discovery ... rests largely on deductive arguments from principles which underlie* not only *science* and deductive metaphysics but also everyday decisions."

On pages 249-250 he argues that

> "the logic of scientific discovery is not inductive; neither does it resemble artistic creation, but is, instead, largely deductive. … Unlike Popper, I think that deduction constitutes the most important element in the process of invention."

The problem with Zahar's proposal is that the deductive character of scientific discovery is based on the metaphysical assumption of the existence of several *philosophical principles* and *meta-principles*, which, as a sort of premises for deductive reasoning, allow for the creation of theories. Such principles are the *Principle of Correspondence*, the *Principle of Identity*, the *Principle of Sufficient Reason*, and the *Principle of Proportionality of Cause and Effect*. Interesting though it is, Zahar's proposal is a metaphysical theory about the deductive nature of scientific discovery. It is a highly speculative hypothesis, far removed from the methodological character both of ampliative inferences in natural observational sciences and of anticipative inferences, like theoretical preduction, in natural theoretical sciences such as mathematical physics.

Contrary to philosophers of science, computational scientists seem to be less reticent when debating the role of deductive reasoning in scientific discovery. For instance, Shrager and Langley (1990, 5-9) emphasize the *deductive law formation*, which

> "produces laws by a second route [the first one is the inductive one, A. R.], starting with a theory and using an explanatory framework to deduce both a law and an explanation of how that law derives from the theory. … For instance, Einstein's theory of general relativity led to an inferred law about the orbit of Mercury."

Indeed, this is a clear instance of how a novel law can be derived in a purely *intra-theorical* way. Further examples are the deduction of Einstein's famous law $E=mc^2$, Maxwell-Boltzmann distribution laws, etc.

The possibility of deductive discovery brings the philosopher of science closer to the idea of *preduction*. As I have advanced elsewhere (see Rivadulla 2010), preduction anticipates still-unknown ideas in physics by the combination, compatible with dimensional analysis, of previously accepted results of different disciplines or theories of mathematical physics. Since these results, which are the premises of the preductive way of reasoning, proceed from different theories, preduction represents *transverse* or *inter-theoretical* deduction. This is what makes it possible to anticipate new ideas in physics.

Is *theory-driven discovery* an antecedent of computational preduction? Simon (1992, 10-11.My italics, A. R.) offers several examples of computational discovery:

> "Suppose we provide BACON with some very broad theoretical concepts: that when substances are mixed together both mass and heat are conserved; and that the law

> describing the temperature equilibrium of such a mixture should be symmetrical in the properties of the components. Then, Black's law of temperature equilibrium can be deduced from these assumptions, in advance of any examination of data, and the data simply used to confirm the law. Without these assumptions, BACON must induce the law from data with the help of rather arduous calculations. Thus, assumptions of conservation and symmetry can be used as heuristics to reduce the search required to find laws."

Similarly, under the assumption that chemical substances are made up of molecules, and molecules of atoms of the elements, and that atoms are conserved in reactions and that volumes of gasses, under constant temperature, are proportional to the numbers of their molecules, etc., the DALTON program

> "deduces the chemical formulas of the molecules involved. For example, on being told that three volumes of hydrogen and one of nitrogen produce two of ammonia, it concludes, correctly, that hydrogen and nitrogen are H_2 and N_2, respectively, and that ammonia has the formula NH_3." (Simon 1992, *ibid.*)

Thus, insofar as theory-driven discovery might involve assumptions from different theories, it could conceivably advance computational preduction.

Computational preduction is automated theoretical preduction. The *computational preduction* hypothesis assumes the possibility of automating preduction mechanisms. It is grounded in the existence, since the 1970s, of computational methods for automating scientific discoveries. Since the applicability of the methods of IA to scientific discovery allows for the automation of inductive and abductive processes, computational preduction would help us to *anticipate* new and previously unknown results in theoretical physics, and would therefore extend scientific creativity.

REFERENCES

Dzeroski, S., *et al.*, 2007, 'Computational Discovery of Scientific Knowledge', in Dzeroski, S. and Todorovski, L. (eds.), *Computational Discovery of Scientific Knowledge. Introduction, Techniques, and Applications in Environmental and Life Sciences*. LNAI 4660, Berlin-Heidelberg, Springer

Langley, P. *et al.*, 1987, *Scientific Discovery. Computational Explorations of the Creative Processes*, Cambridge, MA, The MIT Press

Medawar, P. 1974, 'Hypotheses and Imagination', in Schilpp, A. (ed.), *The Philosophy of Karl Popper*, La Salle, Ill., Open Court, 274-291

Nickles, T., 2008, 'Scientific Discovery', in Psillos, S. & Curd, M. (eds.), *The Routledge Companion to Philosophy of Science*. London, Routledge

Rivadulla, A. 2010, 'Complementary Strategies in Scientific Discovery: Abduction and Preduction', in Bergman, M. *et al.* (eds.), *Ideas in Action: Proceedings of the Applying Peirce Conference*, Nordic Studies in Pragmatism 1, Nordic Pragmatism Network, Helsinki, 264-276

Rivadulla, A. 2012, 'Ciencia computacional de la ciencia. Filosofía del descubrimiento automático', in Fernández Moreno, L. *et al.* (eds.), *Ensayos sobre lógica, lenguaje, mente y ciencia.* Sevilla, Alfar

Shrager, J. and Langley, P., 1990, 'Computational Approaches to Scientific Discovery', in Shrager and Langley, *Computational Models of Scientific Discovery and Theory Decision*, San Mateo, California, Morgan Kaufmann Publishers

Simon, H. A. and Lea, G., 1974, 'Problem Solving and Rule Induction', in Simon, H., *Models of Thought*, New Haven and London, Yale University Press 1979

Simon, H. A., 1992, 'Scientific Discovery as Problem Solving', *International Sudies in the Philosophy of Science* 6, No. 1, 3-14

Simon, H. A., 1995, 'Machine Discovery', *Foundations of Science* 2, 171-200. Reprinted in Zytkow, J. (ed.), *Machine Discovery*, Kluwer Academic Press, Dordrecht 1997

Thagard, P., 1988, *Computational Philosophy of Science*, Cambridge, MA, The MIT Press

Thagard, P. and Nowak, G., 1990, 'The Conceptual Structure of the Geological Revolution', in Shrager and Langley, *op. cit.*

Zahar, E., 1983, 'Logic of Discovery or Psychology of Invention?', *Brit. J. Phil. Sci.* 34, 243-261.