# What can inductive machines suggest about the realism debate

Hilan Bensusan
School of Computing and Cognitive Sciences
University of Sussex
Falmer, Brighton
BN1 9QH
UK
email:hilanb@cogs.susx.ac.uk

January 28, 2014

**Abstract**

This paper attempts to work out a position in the scientific realism debate inspired by the current work on automated induction. Current research on different induction methods suggest that methodology is problem-relative, that the choice of theoretical terms is heavily dependent on the inductive strategy being used and that different inductive procedures produce different empirically equivalent hypotheses. Taken together, these suggestions point to a position that is weaker than scientific realism yet contains the means to provide an explanation for the predictive success of science.

## 1 Introduction

Scientific realism ([5], [6], [26], [29]) is a thesis about scientific theories – the products of scientific induction. One can argue for or against it by appealing to considerations related to science or to features of induction in general. The realism debate is one of the many faces of the philosophical concern about induction. Hume, like many before him, saw how shaky are the foundations of the enterprise of outputting general hypotheses from the meager input of observations. Yet, we usually need to rely on its products. The realist question might be seen as arising when we realise that since not all the products of inductions are to be full-bloodedly believed and we wonder whether there is any reason why science should enjoy a special status.

In the current, naturalist, times, philosophers feel safe to expect aid and inspiration from the natural sciences (see [35], [20]). A less common move is to look at the achievements of the sciences of the artificial[1] such as the study of automated induction in the branch of Artificial Intelligence called Machine Learning[2]. Current research on induction, however, seems to hinge on various issues that have concerned philosophers worried with induction and its products – including science. In this paper I try to examine what this research has to suggest about the realism debate.

Whatever research about machine induction suggests concerning the realism debate, it will be based on general features of induction, rather than on the specific scientific inductions. In the next section I elaborate on how we can say something about the general features of induction. I then proceed to analyse the introduction of theoretical terms and its importance for the realism issue. Inspired by what can be understood about induction from the current Machine Learning work, a position in the realism debate is put forward and some of its advantages scketched. I suggest that it fares better than full-blooded realism and some of the existing forms of antirealism.

## 2 The *inductive continuum*

The study of induction can be carried out either by studying its products or by analysing its processes. The distinction, widely accepted until recently, between the context of discovery and the context of justification has generated a division of labour whereby philosophers would concentrate on the context of justification and therefore on the scientific products of induction – the scientific theories. Philosophers would then study induction by examining the various ways a theory can gain confirmation or support from a body of data. In terms of the scientific realism debate, the question is whether the bonds between the body of data and the theory are strong enough to justify the extension of the ontological status we ascribe to the body of data to the theory.

Recently, due to the naturalist wave, the distinction has become more blurred (see [19]). The naturalist considers that the products are as good

[1]One reason why this move is less common is that the very question of the nature of a science of the artificial is complex and manyfold. The naturalist might want to ask how much of an **empirical** science a science of the artificial is. Of course, the mere fact that empirical methods are applied don't turn a science into an empirical one for these methods can be used only because it might be too complicated (for humans) to use *a priori* methods.

[2]See [37], [22], [28] for an overview of the research in this area.

as the processes that have generated them and good processes are reliable processes[3]. Discovery processes, therefore, become tightly linked to the amount of support a theory enjoys. Naturalists, such as Boyd [4], also consider scientific methods to be as good as the well supported theories they have produced. This is, roughly, how naturalists go about normativity – they would rely on an induction over the known inductive methods relying on inductive products to find out about inductive methods.

In an explicit attempt to provide a naturalist account of epistemic norms, Laudan [25] shows that the naturalist methodologist could rely on historical information to find out by induction which methods are likely to attain a given aim. Building on his reticulated model of scientific inquiry [24], whereby theories promote changes in methods and methods suggest changes in aims and goals, Laudan shows that history of science can be used as data for a methodological induction. The picture that emerges is that induction can be studied inductively, inductive norms being induced from the observation of historical links between methods and aims.

This is where the current research on automated induction enters the scene. The study of inductive methodologies by applying them to a large number of different inductive problems on computers is now possible. Machine Learning is often defined as the study of inductive methodologies or, in the area's terminology, of inductive biases (see [37]) where bias is any basis for choosing one generalisation hypothesis over any other since more than one hypothesis is always empirically adequate to the known data. The concept of bias, therefore, comprises both the representation languages chosen to express the hypotheses and the actual process of generation of a hypothesis. By methodology we mean both the language for induction and the processes carried out in this language.

As an empirical study of induction, Machine Learning proceeds by induction on biases. It aims to study the reliability of different bias such that advise can be given to someone that faces the choice of a methodology for some inductive problem. One easy way to go about biases would be to find a universal bias, appropriate for every induction. Theoretical results, however, show that, in many relevant definitions of appropriateness, there is no such bias (see [16], [40], [36]). Biases promote choices, they can be seen as imposing an ordering on the space of possible hypotheses. The ordering is only as good if the appropriate hypotheses for a given task are on the top. There is no universal ordering for every task. We should therefore rely on

[3]Naturalists are usually commited to a reliabilist epistemology that would define knowledge along the lines of a reliably generated true belief (see [14].

induction about different biases on different environments. Once again, we are left with induction on inductive methods.

The naturalist can find in this Machine Leaning enterprise inspiration for his way to approach induction. He or she can now extend the scope of the study of induction beyond science and towards inductive problems and methods in general. In order to do this, however, a premise is missing. Current research on automated induction does not examine problems as complex as those with which science is frequently involved. No Machine Learning algorithm can induce a theory like Darwinian evolution theory, let alone something like contemporary quantum mechanics. In order for the naturalist to profit from current work on computational induction, it should be granted that there is something like a methodological *inductive continuum* that holds together simple inductive problems and the problems of science. It should be argued that the features of induction that exist in the simpler problems that Machine Learning currently addresses, scale up through the *continuum* towards the inductive tasks of science. There might be various different ways to argue for this *continuum*. One promissing way is to consider induction as a single phenomenon that has general features. In this paper, however, the existence of such *inductive continuum* will not be argued but assumed.

Once the *continuum* is accepted, research on computational induction can inform inductive methodologies and the studies of topics related to induction. Various issues such as the role of background knowledge in hypothesis construction, the importance of using novel evidence to confirm hypotheses and the multiple connections between simplicity and induction (see for instance [2]) are addressed by Machine Learning research. In this paper, I consider issues inductive representation change and their bearings on the realism debate.

## 3 Theoretical terms

One of the lessons to be drawn from current research on machine induction concerns representation changes. Different inductive biases have different ways of approaching problems. Some problems are hard for some biases and easy for others as some methods are more appropriate for some generalisation hypotheses than others. Often, inductive performance can be made easier by changing the representation used for a problem since a given bias uses the problem representation to seek for certain patterns of regularity. By making some correlations between elements of the problem explicit, the bias is changed and the problem is made easier[4]. This attempt to improve representations during the inductive process is often called *constructive induction* ([34], [33], [7]).

Constructive induction can be seen as addressing an issue closely related to the most debated issue concerning the introduction of theoretical terms in scientific theorising (see for instance [38], [18]). Here as there, the questions are about the character and the nature of theoretical terms. The usual way of constructing theoretical terms, mostly inherited from the classical work of Nagel [30] and Hempel [17], involves the dichotomy theoretical/observational. This dichotomy has been exhaustively criticised from various different angles (see [1], [15] and [8]). An alternative definition of theoretical terms can be provided by research in constructive induction: A theoretical term is introduced in a hypothesis when the inductive bias needs it to produce an appropriate hypothesis (according to the bias' standard). This account of the introduction of theoretical terms during inductive processes makes clear that what is relevant is not the definition of a theoretical term in terms of other, non-theoretical ones. In fact, in the case of simple induction, the new terms added in the process are merely relationships between the original predicates and therefore can be explicitly defined in terms of them. What is really relevant is the selection of the relationships between the terms among all possible relationships that could be (explicitly) defined in terms of the original predicates. Whenever the inductive system senses the need for a hypothesis formulated in a richer vocabulary, it selects a relationship that seems appropriate.

[4]To flesh this out, consider the following example. Bias $\beta$ can only represent monomials (conjunctions of $A$, $B$, etc and their negations). The simple problem below is one where the task is to find an expression for $Y$ as a function of $A$ and $B$. The instances for induction have to be one of the following 4.

| $A$ | $B$ | $Y$ |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |

Now, under this representation, $\beta$ cannot generate any predictive hypothesis. However, the problem can also be represented with the help of $C = (-2)A\&B$ :

| $A$ | $B$ | $C$ | $\sum$ | Y |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 1 | 1 |
| 1 | 0 | 0 | 1 | 1 |
| 1 | 1 | -2 | 0 | 0 |

This makes the problem easier for $\beta$. The general problem, of course, is to find the appropriate $C$.

Since theoretical terms are introduced by a given bias, their presence is bias-relative. In fact, different biases need different theoretical terms for the same inductive task. The need for theoretical terms is itself bias-relative. Different biases need theoretical terms for different tasks. This, in its turn, shows that inductive hypotheses are themselves bias-relative and realism about the products of any induction within the scope of these results is challenged by this relativity. Again, if the *inductive continuum* is granted, scientific realism is also challenged.

Hempel's theoretician's dilemma ([18], pg. 186) concludes that theoretical terms are unnecessary for they either define connections among non-theoretical terms or do not serve their purpose and if they define connections among non-theoretical terms they can be replaced by non-theoretical laws. The dilemma concerns the statement of the theories and not their induction or their discovery. In Hempel's terms, theoretical terms can be dispensed with in the formulation of a theory but still be heuristically invaluable[5]. Hempel himself considered theoretical terms to be heuristically useful. However, if we consider the different biases or discovery mechanisms that research on automated induction considers (and grant the *inductive continuum*), theoretical terms are useful only for some inductive biases. In general, we can say that theoretical terms are needed whenever the bias is not appropriate to the problem at hand. By measuring the bias of a learning system [3], it is possible to show that the bias is weakened by the introduction of theoretical terms. A weaker bias means a larger space of hypotheses to be considered and/or a softer preference for some hypotheses.

The bias-relativity of the need for theoretical terms doesn't mean that there is a bias that needs no theoretical terms for whatever problem. This seems highly implausible given the theoretical results mentioned above. It

[5]The message of the theoretician's dilemma has been partly inspired by Craig's result [9]. Craig showed that if a theory can be recursively axiomatised with theoretical terms present in the axioms, it can be equally recursively axiomatised without them (using observational terms as the only non-logical vocabulary). If one says that the purpose of theoretical terms is to state connections between observable phenomena in a recursively axiomatised theory, then they are not needed. If, however, we move from recursive axiomatisability to the more restricted requirement of finite axiomatisability, theoretical terms are needed (see [10]). Therefore, a scientific theory - or perhaps the whole scientific corpus if we assume that it can be axiomatised - can be expressed without theoretical terms if we are prepared to admit that a recursive (non-finite) axiomatisation is adequate. If we are prepared to do so, the conclusion is that science can be stated (at least in principle) without theoretical terms. But, as Hempel said [18], Craig didn't show that theoretical terms are heuristically dispensable. The heuristic need for theoretical terms began to be challenged by a result [12] coming from the identification in the limit learning tradition [31], [13].

doesn't entail either that there is a bias that needs no theoretical terms for whatever scientific problem. But it does entail an argument against the reference of theoretical terms: no matter how predictively successful a theory is, its theoretical terms are heuristically useful for the bias that has required it. Two different biases might come up with equally empirically adequate hypotheses using two disjoint (and non-intertranslatable) classes of theoretical terms. It is as if the study of biases provided by current Machine Learning has enabled us to step backwards and consider the relativity of our heuristic requirements. We can now see our scientific or ordinary inductive machinery as one among others and its products as relative to our ways of proceeding inductively.

## 4 Surrealism

The above considerations suggest a position in the realism debate. Scientific realism is the claim that theoretical terms of mature and well-confirmed theories refer to bits of reality and are approximately true. The arguments in favour of such a position lie in confirmation grounds. It is often pointed out that the predictive success of mature science could not be accounted for without the postulation of the approximate truth of its theories (understanding approximate truth as something that involves the reference of the terms used to state the theories). This is often called the non-miracle argument for it is said that only a miracle could explain the predictive success of these theories if they are not approximately true and, in particular, if the central, or most, or some theoretical terms do no refer (see formulations in [32], [29], [6], [39]). The non-miracle argument claims therefore that realism is the only possible explanation of the successful novel predictions of well-confirmed mature scientific theories.

Although there are a number of problems related to what counts as a successful new and independent prediction, what is approximate truth and the nature of the sort of inference to the best explanation used in the argument, I shall consider that the non-miracle argument is basically compelling. The argument shows that, when compared to anti-realists, realists fare better concerning the explanation of the success of science. However, the considerations about the bias-relativity of the need for theoretical terms suggest that a realist position is too strong to be justifiable. Theoretical terms are partially determined by the world, by the inductive problems but also partially determined by our inductive biases. We need something weaker than realism and yet strong enough to provide an explanation of the predictive success of science.

A non-realist explanation of the success of science has been sought by antirealists as a way to crack down the non-miracle argument (see for instance [39], [11], [23]). The realist Leplin [27] suggested that an as-if-like explanation could be put forward and calls it *surrealism.* Surrealism is not commited to the belief that science represents the deep structure of the world: the world is as if our mature, well-confirmed theories were true. In other words, the empirical consequences of our best theories are as a matter of fact the same as the empirical consequences of the deep structures of the world. Leplin argues against this position by attempting to show that surrealism is either non-explanatory or collapses into full-bloded realism.

In [21], Kukla attempts to defend surrealism against Leplin's arguments. He distinguishes between what he calls *weak surrealism* and *strong surrealism.* Weak surrealism is not commited to the existence of deep structures of the world. It asserts that there is a natural law that asserts that the observable world behaves as if the theories are true. If $E$ is discovered to be an empirical consequence of the best theories, the weak surrealist will immediately believe $E$. According to the weak surrealist, this law connecting the observable world to our mature, well-confirmed theories explains the success of science.

Leplin, however, argues that realism provides an explanation for the surrealist *explanans* of success that would otherwise be relying on the unexplained brute fact that such a natural law holds. Realism would explain this law by appealing to the approximate truth of the theories and therefore to the reference of some or most of the theoretical terms used to express the theory. If we assume that there should be no limits to the demand for explanation, we should then take this path. However, this move is prevented by the bias-relativity considerations made above. If the *inductive continuum* holds, our mature, well-confirmed scientific theories are a product of an inductive bias and approximate truth and reference compose no available *explanans.* We can then ask whether these considerations can provide an alternative explanation to the law that connects the predictions of our theories to the theories themselves. In general, we can ask whether these considerations can spell out an alternative to weak surrealism.

If we want to keep the explanatory power of weak surrealism, we should preserve the natural law that does the explanatory work. I submit that a good alternative is close to the Kantian position about human knowledge. We can call this position *transcendental surrealism.* We can postulate, rather in a Kantian way, that the link between our theories and our predictions is given by the appropriateness of our theories for us. In our terms, the transcendental subject that builds knowledge from sensorial impressions is an inductive bias and we could then talk of a number of different transcendental subjects[6]. According to transcendental surrealism, the success of science can be explained by the adequacy (or approximate adequacy) of our theories to our inductive machinery. Instead of talking about truth, we talk of adequacy given a bias. An immediate consequence of this explanation is that we need nothing like a property of reference to spell out our notion of adequacy[7].

The transcendental surrealist explanation of success postulates that our mature, well-confirmed theories have a special and strong relation to our inductive biases. It has explanatory power from this connection between theories and biases that, as for the weak surrealist, is a nomological connection. Somehow, however, this nomological connection is also a transcendental one since it connects biases and hypotheses. We should not loose sight of the strength of this position. Its explanatory power comes from the postulation that there is something special about our theories although the special thing has nothing to do with truth (at least in the correspondence construal of truth). Transcendental surrealism is therefore stronger than constructive empiricism [39] since it acknowledges that there is something special about our mature, well-confirmed theories apart from their empirical adequacy. However, it has no problems with thesis such as the unavoidable underdetermination of incompatible theories by data since more than one theory can hold this special adequacy relation to our biases. As a general advantage over instrumentalism and constructive empiricism, transcendental surrealism offers a non-realist explanation of the success of science.

It is clear that our position is also weaker than full-blooded scientific realism. According to transcendental surrealism, theories and hypotheses are relative to both the way the world is and the way a inductive machinery works. A number of incompatible theories might be adequate for both the world and a given inductive machinery. Also, we cannot rule out that other, different inductive biases would generate other, incompatible sciences. Although this position needs a more complete analysis, it looks like references shall melt like clocks in Dali's paintings.

[6]Transcendental surrealism makes the study of computational induction a kind of transcendental subject whereby the possibilities of different transcendental subjects are analysed.

[7]Incidentally, this protects transcendental surrealism from one of the most damaging arguments against scientific realism: a Laudanian gambit whereby historical cases of success without reference undermine the attempt to strongly link the two together ([24], chap. 5).

## References


[1] P. Achinstein. The problem of theoretical terms. *American Philosophical Quarterly*, 2(3), 1965.

[2] H. Bensusan. God doesn't always shave with Occam's Razor – learning when and how to prune. In C. Nédellec and C. Rouveirol, editors, *Proceedigs of the 10th European Conference on Machine Learning*, pages 119–124, Berlin, 1998. Springler.

[3] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, 36(4):929–965, 1989.

[4] R. Boyd. The logician's dilemma. *Erkenntnis*, 22:197–152, 1985.

[5] R. Boyd. What realism implies and what it does not. *Dialectica*, 43:5–29, 1989.

[6] R. Boyd. Realism, approximate truth and philosophical method. In *Scientific Theories, Minnesota Studies in the Philosophy of Science, 14*, pages 355–391. University of Minnesota Press, Minneapolis, MN, USA, 1990.

[7] J. Callan, T. Fawcett, C. Matheus, R. Michalski, M. Pazzani, L. Rendell, and R. Sutton, editors. *ML-COLT '94 Workshop on Constructive Induction and Change of Representation*, 1994.

[8] P. M. Churchland. *Scientific Realism and the Plasticity of Mind.* Cambridge University Press, Cambridge, MA, USA, 1979.

[9] W. Craig. On axiomatizability within a system. *The journal of Symbolic logic*, 18(1):30–32, 1953.

[10] W. Craig and R.S. Vaught. Finite axiomatizability using additional predicates. *Journal of Symbolic Logic*, 23:289–308, 1958.

[11] A. Fine. Unnatural attitudes: Realist and instrumentalist atachments to science. *Mind*, 95:149–179, 1986.

[12] H Gaifman, D. Osherson, and S. Weinstein. A reason for theoretical terms. *Erkenntnis*, 32:149–159, 1990.

[13] C. Glymour. Inductive inference in the limit. *Erkenntnis*, 22:23–31, 1985.

[14] A. Goldman. What is justified belief? In H. Kornblith, editor, *Naturalising Epistemology*, pages 105–130. MIT Press, Cambridge, Mass., USA, 1994.

[15] N. R. Hanson. *Perception and Discovery.* Freeman and Cooper Co, USA, 1969.

[16] D. Haussler. Quantifying inductive bias: AI learning algorithms and Valiant's learning framework. *Artificial Intelligence*, 32(2):177–222, 1988.

[17] C. G. Hempel. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science.* Free Press, New York, NY, USA, 1965.

[18] C. G. Hempel. The theoretician's dilemma. In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, pages 173–226. Free Press, New York, NY, USA, 1965.

[19] A. Kantorovich. Philosophy of science: From justification to explanation. *The British Journal for the Philosophy of Science*, 39:469–494, 1988.

[20] H. Kornblith, editor. *Epistemology Naturalised.* MIT Press, Cambridge, MA, USA, 1994.

[21] A. Kukla. Antirealist explanations of the success of science. *Philosophy of Science*, 63:S298–S305, 1996.

[22] P. Langley. *Elements of Machine Learning.* Morgan Kaufmann, 1996.

[23] L. Laudan. Explaining the success of science: beyond epistemic realism and relativism. In J. T. Crushing, C. F. Delaney, and G. M. Gutting, editors, *Science and Reality*, pages 83–105. University of Notre Dame Press, Notre Dame, IL, USA, 1984.

[24] L. Laudan. *Science and Values.* UC Press, Berkeley, CA, USA, 1984.

[25] L. Laudan. Progress or rationality? the prospects for normative naturalism. *American Philosophical Quarterly*, 24:19–31, 1987.

[26] J. Leplin. *Scientific Realism.* University of California Press, Berkeley, CA, USA, 1984.

[27] J. Leplin. Surrealism. *Mind*, 96:519–524, 1987.

[28] T. Mitchell. *Machine Learning.* McGraw Hill, 1997.

[29] A. Musgrave. The ultimate argument for scientific realism. In R. Nola, editor, *Relativism and Realism in Sciences.* Kluwer Academic Press, Dordrecht, 1988.

[30] E. Nagel. *The structure of Science.* Hackett Publishing Co, Indianapolis, IN, USA, 1960.

[31] D. Osherson, M. Stob, and S. Weinstein. *Systems that Learn.* MIT Press, Cambridge, MA, USA, 1986.

[32] H. Putnam. *Mathematics, matter, and method.* Cambridge University Press, Cambridge, MA, USA, 1975.

[33] L. Rendell and H. Cho. Empirical learning as a function of concept character. *Machine Learning*, 5:267–298, 1990.

[34] L Rendell and R. Seshu. Learning hard concepts through constructive induction: Framework and rationale. *Computational Intelligence*, 9:247–270, 1990.

[35] A. Rosenberg. A field guide to recent species of naturalism. *British Journal for the Philosophe of Science*, 47:1–29, 1996.

[36] C. Schaffer. A conservation law for generalization performance. In *Proceedings of the Eleventh Conference on Machine Learning*, San Mateo, CA, USA, 1994. Morgan Kaufmann.

[37] J. Shavlik and T. Dietterich. *Readings in machine learning.* Morgan Kaufman, 1990.

[38] F. Suppe. The search for philosophic understanding of scientific theories. In F. Suppe, editor, *The structure of scientific theories*, pages 3–241. University of Illinois Press, 1977,.

[39] B. van Fraassen. *The Scientific Image.* Oxford University Press, Oxford, UK, 1980.

[40] D Wolpert. The lack of a priori distinctions between learning algorithms. *Neural Computation*, 8:1341–1390, 1996.