

# Induction and Physical Theory Formation As Well As Universal Computation by Machine Learning

---

Alexander Svozil<sup>1</sup> and Karl Svozil<sup>2</sup>

<sup>1</sup> Theory and Applications of Algorithms, Faculty of Computer Science,  
University of Vienna Währinger Straße 29, A-1090 Vienna, Austria

<sup>2</sup> Institute for Theoretical Physics, Vienna University of Technology,  
Wiedner Hauptstraße 8-10/136, A-1040 Vienna, Austria

## 1. Algorithmic Induction

There appears to be at least two approaches towards induction. The first route is by intuition and ingenuity. This route has been successfully pursued by geniuses and gifted individuals. A typical representative of this approach to knowledge is Ramanujan who seemed to have attributed his revelations to a Hindu Goddess [1]. In western thought, this is often more secularly referred to as Platonism. Gödel appeared to have held the opinion that our minds have access to the truth, which can be discovered through personal insights – perhaps even beyond the bounds of universal computability – in particular, the idea that minds are no (Turing) machines [2, p. 216]. As successful as these narratives may have been, they remain anecdotal and cannot be generalized.

When it comes to *ad hoc* revelations of individuals, there may also be psychological issues. These have been described by Freud [3], as pointing to the dangers caused by “temptations to project, what [the analyst] in dull self-perception recognizes as the peculiarities of his own personality, as generally valid theory into science.” A similar warning comes from Jaynes’ “Mind Projection Fallacy” [4, 5], pointing out that “*we are all under an ego-driven temptation to project our private thoughts out onto the real world, by supposing that the creations of one’s own imagination are real*

---

\*Corresponding author: alexander.svozil@nivio.ac.at, svozil@tuwien.ac.at

*properties of Nature, or that one's own ignorance signifies some kind of indecision on the part of Nature."*

A second, computational, approach could be conceived in the spirit of Turing [6]. In this line of thought, it is possible to obtain knowledge about a system by mechanical, algorithmic procedures; such as a deterministic agent "*provided with paper, pencil and rubber, and subject to strict discipline [carrying out a set of rules of procedure written down]*" [7, p. 34]. Recently one of the more promising methods to algorithmic induction has been machine learning [8] which will be pursued in this chapter. There is even a form of collective intuition – the so-called *swarm intelligence* [9, 10, 11] – that can solve many logistic problems effectively.

Two *caveats* should be stated upfront. First, that the general induction problem is – in recursion theoretic terms, the rule inference problem – unsolvable with respect and relative to universal computational capacities [12, 13, 14, 15]. So, in certain (even constructible) situations machine learning, like all other algorithmic induction strategies, provably fails. But that does not exclude heuristic methods of induction, such as machine learning applied to physical phenomena. Secondly, theoretical constructions cannot be expected to faithfully represent the "laws underlying" a phenomenology. As Lakatos [16] points out, the progressiveness and degeneracy of research programs are transient; and without a recognizable coherent conceptual convergence. Therefore, the "explanations" and theoretical models generated by machine learning present knowledge and explanations which cannot claim to have any ontologic (only epistemic) relevance – they are means relative to the respective methods employed.

Whereas machine learning has already been applied to very specific problems in high energy [17] and solid-state physics [18], we would like to propose this as an extremely general method of theory formation and induction.

We are applying the obtained machine representations to predict a simulacrum – we are, in particular, interested in universal Turing computation; at least until a finite amount of computational space and time [19].

## 2. Linear Models of Inertial Motion

In what follows a *linear regression* model (Sect. 5.1.4) [8], will be used. Thereby we shall, if not stated otherwise, explicitly closely follow the notation of Mermin's book on *Quantum Computer Science* [20].

Suppose, for the sake of demonstration, a one-dimensional physical system of a particle in inertial motion. Suppose further that it has been (approximately) measured already at  $n \geq 2$  positions  $x_1, \dots, x_n$  at times  $t_1, \dots, t_n$ , respectively. The goal is to find a general algorithm which predicts its location at an arbitrary time  $T$ .

In what follows the respective positions and times are (not necessarily successively) arranged as  $n$ -tuples; that is, as a finite ordered list of elements, and interpreted as  $(n \times 1)$ -matrices

$$\begin{aligned} |x\rangle &\equiv (x_{i_1}, \dots, x_{i_n})^T \\ |t\rangle &\equiv (t_{i_1}, \dots, t_{i_n})^T \end{aligned} \quad (1)$$

where superscript  $T$  indicates transposition, and  $i_1, \dots, i_n$  are arbitrary permutations of  $1, \dots, n$ . That is, it is not necessary to order the events temporally; actually, they can “run backward” or be randomly arranged [21].

A linear regression *Ansatz* is used to find a linear model for the prediction of some unknown observable, given some of the anecdotal instances of its performance. More formally, let  $y$  be an arbitrary observable which depends on  $n$  parameters  $x_1, \dots, x_n$  by linear means; that is, by

$$y = \sum_{i=1}^n x_i r_i = \langle \mathbf{x} | \mathbf{r} \rangle \quad (2)$$

where  $\langle \mathbf{x} | = (\mathbf{x})^T$  is the transpose of the vector  $|\mathbf{x}\rangle$ , the tuple

$$|\mathbf{r}\rangle = (r_1, \dots, r_n)^T \quad (3)$$

contains the unknown weights of the approximation – the “theory,” if you like – and  $\langle \mathbf{a} | \mathbf{b} \rangle = \sum_i a_i b_i$  stands for the Euclidean scalar product of the tuples interpreted as (dual) vectors in  $n$ -dimensional (dual) vector space  $\mathbb{R}^n$ .

Given are  $m$  known instances of (2); that is, suppose  $m$  pairs are  $(z_j, |\mathbf{x}_j\rangle)$  known. These data can be bundled into an  $m$ -tuple

$$|\mathbf{z}\rangle \equiv (z_{j_1}, \dots, z_{j_m})^T \quad (4)$$

and an  $(m \times n)$ -matrix

$$\mathbf{X} \equiv \begin{pmatrix} x_{j_1 i_1} & \dots & x_{j_1 i_n} \\ \vdots & \vdots & \vdots \\ x_{j_m i_1} & \dots & x_{j_m i_n} \end{pmatrix} \quad (5)$$

where  $j_1, \dots, j_m$  are arbitrary permutations of  $1, \dots, m$ , and the matrix rows are just the vectors  $|\mathbf{x}_{j_k}\rangle \equiv (x_{j_k i_1}, \dots, x_{j_k i_n})^T$ .

The task is to compute a “good” estimate of  $|\mathbf{r}\rangle$ ; that is, an estimate of  $|\mathbf{r}\rangle$  which allows for an “optimal” computation of the prediction  $y$ .

Suppose that a good way to measure the performance of the prediction from some particular definite but unknown  $|\mathbf{r}\rangle$  with respect to the  $m$  given data  $(z_j, |\mathbf{x}_j\rangle)$  is by the mean squared error (MSE)

$$\begin{aligned}
 \frac{1}{m} \|\mathbf{y} - |\mathbf{z}\rangle\|^2 &= \frac{1}{m} \|\mathbf{X}|\mathbf{r}\rangle - |\mathbf{z}\rangle\|^2 \\
 &= \frac{1}{m} (\mathbf{X}|\mathbf{r}\rangle - |\mathbf{z}\rangle)^T (\mathbf{X}|\mathbf{r}\rangle - |\mathbf{z}\rangle) \\
 &= \frac{1}{m} (\langle \mathbf{r} | \mathbf{X}^T - \langle \mathbf{z} |) (\mathbf{X}|\mathbf{r}\rangle - |\mathbf{z}\rangle) \\
 &= \frac{1}{m} (\langle \mathbf{r} | \mathbf{X}^T \mathbf{X} | \mathbf{r} \rangle - \langle \mathbf{z} | \mathbf{X} | \mathbf{r} \rangle - \langle \mathbf{r} | \mathbf{X}^T | \mathbf{z} \rangle + \langle \mathbf{z} | \mathbf{z} \rangle) \\
 &= \frac{1}{m} [\langle \mathbf{r} | \mathbf{X}^T \mathbf{X} | \mathbf{r} \rangle - \langle \mathbf{z} | (\langle \mathbf{r} | \mathbf{X}^T)^T - \langle \mathbf{r} | \mathbf{X}^T | \mathbf{z} \rangle + \langle \mathbf{z} | \mathbf{z} \rangle] \\
 &= \frac{1}{m} \left\{ \langle \mathbf{r} | \mathbf{X}^T \mathbf{X} | \mathbf{r} \rangle - [(\langle \mathbf{r} | \mathbf{X}^T)^T | \mathbf{z} \rangle - \langle \mathbf{r} | \mathbf{X}^T | \mathbf{z} \rangle + \langle \mathbf{z} | \mathbf{z} \rangle] \right\} \\
 &= \frac{1}{m} (\langle \mathbf{r} | \mathbf{X}^T \mathbf{X} | \mathbf{r} \rangle - 2 \langle \mathbf{r} | \mathbf{X}^T | \mathbf{z} \rangle + \langle \mathbf{z} | \mathbf{z} \rangle) . \tag{6}
 \end{aligned}$$

In order to minimize the mean squared error (6) with respect to variations of  $|\mathbf{r}\rangle$  one obtains a condition for “the linear theory”  $|\mathbf{y}\rangle$  by setting its derivatives (its gradient) to zero; that is

$$\partial_{|\mathbf{r}\rangle} \text{MSE} = 0. \tag{7}$$

A lengthy but straightforward computation yields

$$\begin{aligned}
 \frac{\partial}{\partial r_i} (r_j \mathbf{X}_{jk}^T \mathbf{X}_{kl} r_l - 2 r_j \mathbf{X}_{jk}^T z_k + z_j z_j) \\
 &= \delta_{ij} \mathbf{X}_{jk}^T \mathbf{X}_{kl} r_l + r_j \mathbf{X}_{jk}^T \mathbf{X}_{kl} \delta_{il} - 2 \delta_{ij} \mathbf{X}_{jk}^T z_k \\
 &= \mathbf{X}_{ik}^T \mathbf{X}_{kl} r_l + r_j \mathbf{X}_{jk}^T \mathbf{X}_{ki} - 2 \mathbf{X}_{ik}^T z_k \\
 &= \mathbf{X}_{ik}^T \mathbf{X}_{kl} r_l + \mathbf{X}_{ik}^T \mathbf{X}_{kj} r_j - 2 \mathbf{X}_{ik}^T z_k \\
 &= 2 \mathbf{X}_{ik}^T \mathbf{X}_{kj} r_j - 2 \mathbf{X}_{ik}^T z_k \\
 &= 2(\mathbf{X}^T \mathbf{X} | \mathbf{r} \rangle - \mathbf{X}^T | \mathbf{z} \rangle) = 0 \tag{8}
 \end{aligned}$$

and finally, upon multiplication with  $(\mathbf{X}^T \mathbf{X})^{-1}$  from the left,

$$|\mathbf{r}\rangle = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T | \mathbf{z} \rangle. \tag{9}$$

A short plausibility check for  $n = m = 1$  yields the linear dependency  $|\mathbf{z}\rangle = \mathbf{X} | \mathbf{r} \rangle$ .

Coming back to the one-dimensional physical system of a particle in inertial motion, we could characterize inertial motion in machine learning and linear regression terms by the requirement that the Ansatz (2) is “good” in the sense that predictions can be made to a “sufficient degree” (a term which is arbitrary, subjective and thus conventional); that is, within a pre-defined error.

If the particle does not pass through the origin, it might be necessary to augment Eq. (2) with an affine term  $d$ , which can be absorbed into  $|\mathbf{x}'\rangle = (x'_1, \dots, x'_{n+1}) = (x_1, \dots, x_n, 1)^T$  and  $|\mathbf{r}'\rangle = (r'_1, \dots, r'_{n+1}) = (r_1, \dots, r_n, d)^T$  such that

$$y = \sum_{i=1}^n x_i r_i = \langle \mathbf{x} | \mathbf{r} \rangle + d = \sum_{i=1}^{n+1} x'_i r'_i \quad (10)$$

### 3. Nonlinear Models of Noninertial Motion

The linear Ansatz (2) fails for noninertial motion. One possible way to cope with nonlinearities would be to introduce extra dimensions corresponding to nonlinear terms, such as  $x^l$  for  $2 \leq l \leq d < \infty$ ; with the consequence that the dimensionality of the parameter space increases.

To cope with nonlinear phenomena, *deep forward networks* have been used in machine learning [8, Chapt. 6]. This strategy of deep learning invokes intermediate *hidden theoretical layers* of description which communicate with each other. For the sake of an example, suppose there are two functions  $g$  and  $h$ , connected in a chain by functional substitution, such that  $f(x) = h(g(x))$ . The length of the chain is identified with the *depth* of the model – in this case two.  $g$  is the first layer of the model. The final layer – in this case  $h$  – is called the *output layer*.

Unfortunately, the linear regression Ansatz (10) for  $g$  and  $h$  would effectively be linear again. Therefore, to model a nonlinear phenomenology, a nonlinear Ansatz for at least one layer has to be implemented. Such networks are capable of approximating any Borel measurable function (and its derivative, even if it is a generalized function) from one finite-dimensional space to another [22, 23, 24].

### 4. Simulation of Universal Turing Machines by Deep Forward Networks

It could be suspected that, even though for all practical purposes [25] the methods and techniques discussed so far are “good,” the task of finding “better and better” approximations or even total correspondence might, at least in some cases, turn out to be “difficult” if not outrightly impossible. Because suppose, for the sake of a *reductio ad absurdum* – more precisely, a reduction to the rule inference and halting problems

– that such learning of the exact behavior would be computable. Such a suspicion might be tempting, considering the approximate “solution” of the general rule inference problem [12, 13, 14, 15] suggested earlier in the limit of infinite precision. Alas, any such unbounded computation, as long as it needs to be finite, would run into the problem that no computable rate of convergence can be given – very much like the Busy Beaver function [26] or Chaitin’s halting probability  $\Omega$  [27, 28]. Formally, any such claim can be reduced to the rule inference and the halting problem of universal computers. The situation is not dissimilar to series solutions of the  $n$ -body problem [29], which may converge “very slowly” (indeed, intractably slow in numerical work [30]); but if the system encodes a universal computer, they cannot converge in general due to the reduction to the halting problem [31].

However, one could (courageously) “invert” or transform these objections and deficiencies into virtues [32, 33] and argue that, for all practical purposes and in many relevant instantiations, machine learning, and, in particular, deep forward networks, may turn out to be effective in the simulation of universal computers such as a universal Turing machine. Thereby the criterion is not to obtain an exact correspondence; rather the result of the simulation should be “good enough” to justify its use.

For the sake of an example, think of the typical physical estimate of a quantity in terms of *orders of magnitude*: very often, the applicability of a suggestion or technique does not depend on the exact observable value it generates but rather on the order of its impact. This is not dissimilar to certain quantum advantages: for instance, the Deutsch algorithm identifies the parity or (non)constancy of a binary function of a single bit without identifying the actual function [20]; thereby rendering a partitioning of the set of all such individual functions.

How could one imagine training a deep forward network to simulate a universal computer? Fairly simple: let it access the input-output behavior of actual “exact” devices with von Neuman architecture. After “lots of” training, the network should be able to emulate the performance of this architecture within “reasonable precision.” That is, it won’t be able to give the exact value of, say, an algebraic operation like  $n + m$  for “large” (for physical realizability) numbers  $n$  and  $m$ . But it might be able to output some estimates which are “close to” (for the applicability) the exact result.

One could also say that, in this scheme, the deep forward network acts as an oracle with respect to the universal computer. And although the training of it may take some time because of the sheer training volume, as well as the conceivable computational complexity of the individual input-output functions involved, eventually, the trained network is not bound by these restrictions and can reach an (approximate) result quite fast; such that the computation time it takes for any simulation

is uniformly bounded from above. In particular, the halting problem can be said to be “for all practical purposes (FAPP) solvable” by such oracles; but, of course, no guarantee of validity or total precision can be given. Inadvertently one cannot exclude instances in which the deep forward network predicts halting whereas the universal computer it simulates does not halt, and vice versa.

One may also ask: where exactly is the physical resource rendering the computational capacity of such “universal” deep forward networks located? There has to be some formal symbolic encoding in terms of physical components or entities making the simulation feasible and effective. One rather straightforward way to answer this is in terms of the *connections* or *correlations* among the nodes involved: if they are modeled as continuous formal entities such as real or complex numbers then the capacity of even a finite such configuration to store information is unbounded.

In more pragmatic, practical terms the non-exact but effective simulation of general (universal) computations may also present a way to circumvent the stall in Moore’s law which can be observed already for a couple of years. Currently, because of physical restrictions on circuit and switch designs, most improvements in performance are due to parallelization rather than miniaturization. Eventually, the switching time of electronic devices is restricted by fundamental limits from below on resistance; in particular the von Klitzing constant  $R_K$ .

## 5. Discussion

One objection for applying machine learning algorithms to physical theory creation or simulation of universal computers might be that the resulting representations lack any sort of “meaning;” that is, these representations amount to pure syntax devoid of any conceptual semantics. But if conceptual semantics is omitted, there can be no true “understanding” of the “physics behind” the phenomena, or the computational processes yielding those estimates.

One may counter this criticism by noticing that, first, underlying such objections is the premise that something can actually be discovered or revealed. This realistic ontology is by far from non-trivial and is heavily debated [34, 35, 36]. If, for example, the phenomena emerges from primordial chaos, such as in Greek mythology and cosmology,  $\chi\alpha\omicron\varsigma$ , then any “meaning” one might present and “discover” ultimately remains a (pragmatic) narrative or a mathematical abstraction such as Ramsey theory at best: for any data, there cannot be no correlations – regardless of the origin or type of empirical data, there always has to be some, maybe spurious [37, 38], regularity or coincidences or properties.

How can it be excluded that the laws of physics mean nothing but yet undiscovered consequences of Ramsey theory?

Second, as has already been pointed out, historic evidence seems to suggest that successive physical conceptual models (say, of gravity) are not continuously evolving; but that they are disruptive and dissimilar [39, 16]: they lack conceptual convergence. One may even go so far as to suggest that, in any case, theories are (more or less [40]) successful belief systems; very much like Greek mythology [41].

Third, also the present perception of the quantum mechanical formalism includes, among other inclinations, the position that no interpretation is necessary [42]; that indeed, interpretation is even dangerous and detrimental for the researcher [43, p. 129]; or that, at the very least, there are no issues with respect to interpretation [44].

Nevertheless, it might be quite amusing to study toy universes capable of universal computation, such as Conway's *game of life*, via intrinsic, embedded, machine learning algorithms. It cannot be excluded that these kinds of algorithmic agents "come up" with the "right rules;" that is those rules which define the toy mini-universe. It can be expected that if a machine learning algorithm performs excellently on particular problems then it necessarily has a degraded performance on the set of all remaining problems [45, 46].

In any case, the manner in which physical theories are created and invented by human individuals is not dissimilar from machine learning. And machine learning might become of great practical utility for the simulation of (universal) computations.

## Acknowledgments

This work was supported in part by the European Union, Research Executive Agency (REA), Marie Curie FP7-PEOPLE-2010-IRSES-269151-RANPHYS grant. Responsibility for the information and views expressed in this article lies entirely with the authors. The authors declare no conflict of interest.

## References

1. R. Kanigel. *The Man Who Knew Infinity: A Life of the Genius Ramanujan*, 5th ed. Washington Square Press (1991).
2. G. Kreisel. Kurt Gödel. 28 April 1906-14 January 1978. Biographical memoirs of Fellows of the Royal Society 26, 148-224, corrections Ibid. 27, 697; *ibid.* 28, 718 (1980).
3. S. Freud. Ratschläge für den Arzt bei der psychoanalytischen Behandlung. In: *Gesammelte Werke. Chronologisch geordnet. Achter Band. Werke aus*



- den Jahren 1909–1913, edited by Anna Freud, E. Bibring, W. Hoffer, E. Kris and O. Isakower, Fischer, Frankfurt am Main, 1912, 1999, pp. 376–387 (1999).
4. E.T. Jaynes. Clearing up mysteries – the original goal. *In*: Maximum-Entropy and Bayesian Methods: Proceedings of the 8th Maximum Entropy Workshop, held on August 1-5, 1988, in St. John's College, Cambridge, England, edited by John Skilling, Kluwer, Dordrecht, pp. 1–28 (1989).
  5. E.T. Jaynes. Probability in quantum theory. *In*: Complexity, Entropy, and the Physics of Information: Proceedings of the 1988 Workshop on Complexity, Entropy, and the Physics of Information, held May-June, 1989, in Santa Fe, New Mexico, edited by Wojciech Hubert Zurek Addison-Wesley, Reading, MA, pp. 381–404 (1990).
  6. A.M. Turing. Intelligent machinery, a heretical theory. *Philosophia Mathematica* 4, 256–260 (1996).
  7. A.M. Turing. Intelligent machinery. *In*: Cybernetics. Key Papers, edited by C.R. Evans and A.D.J. Robertson. Butterworths, London, pp. 27–52 (1968).
  8. I. Goodfellow, Y. Bengio and A. Courville. Deep Learning (2016).
  9. E. Bonabeau, M. Dorigo and G. Theraulaz. Swarm Intelligence: From Natural to Artificial Systems. Santa Fe Institute Studies on the Sciences of Complexity. Oxford University Press, New York, NY (1999).
  10. J. Kennedy, R.C. Eberhart and Y. Shi. Swarm Intelligence. The Morgan Kaufmann Series in Evolutionary Computation. Morgan Kaufmann, Academic Press. Elsevier, San Francisco, San Diego, CA (2001).
  11. I. Zelinka and G. Chen. Evolutionary Algorithms, Swarm Dynamics and Complex Networks. Methodology, Perspectives and Implementation, Emergence, Complexity and Computation. Vol. 26. Springer-Verlag, Berlin, Heidelberg (2018).
  12. M.E. Gold. Language identification in the limit. *Information and Control*. 10, 447–474 (1967).
  13. L. Blum and M. Blum. Toward a mathematical theory of inductive inference. *Information and Control* 28, 125–155 (1975).
  14. Dana Angluin and Carl H. Smith. Inductive inference: Theory and methods. *ACM Computing Surveys* 15, 237–269 (1983).
  15. L.d.M. Adleman and M. Blum. Inductive inference and unsolvability. *The Journal of Symbolic Logic* 56, 891–900 (1991).
  16. I. Lakatos. Philosophical Papers. 1. The Methodology of Scientific Research Programmes. Cambridge University Press, Cambridge (1978).
  17. Kaggle Team. The HiggsML challenge: When high energy physics meets machine learning. May-September 2014, accessed August 31, 2016.
  18. L.-F. Arsenault, O. Anatole von Lilienfeld and A.J. Millis. Machine learning for many-body physics: Efficient solution of dynamical mean-field theory. arXiv: 1506.08858 (2015).
  19. R.O. Gandy. Limitations to mathematical knowledge. *In*: Logic colloquium '80, edited by D. van Dalen, D. Lascar and J. Smiley. North Holland, Amsterdam, papers intended for the European Summer Meeting of the Association for Symbolic Logic, pp. 129–146 (1982).

20. D.N. Mermin. Quantum Computer Science. Cambridge University Press, Cambridge (2007).
21. G. Egan. Permutation City (1994). Accessed January 4, 2017.
22. K. Hornik, M. Stinchcombe and H. White. Multilayer feed-forward networks are universal approximators. *Neural Networks* 2, 359–366 (1989).
23. K. Hornik, M. Stinchcombe and H. White. Universal approximation of an unknown function and its derivatives using multilayer feedforward networks. *Neural Networks* 3, 551–560 (1990).
24. K. Hornik. Approximation capabilities of multilayer feed forward networks. *Neural Networks* 4, 251–257 (1991).
25. J. S. Bell. Against ‘measurement’. *Physics World* 3, 33–41 (1990).
26. G.J. Chaitin. Computing the busy beaver function. *In*: Thomas M. Cover and B. Gopinath (eds), *Open Problems in Communication and Computation*. p. 108, Springer, New York (1987).
27. C.S. Calude, M.J. Dinneen and Chi-Kou Shu. Computing a glimpse of randomness. *Experimental Mathematics* 11, 361–370 (2002), arXiv:nlin/0112022.
28. C.S. Calude and M.J. Dinneen. Exact approximations of omega numbers. *International Journal of Bifurcation and Chaos* 17, 1937–1954 (2007), CDMTCS report series 293.
29. Qui Dong Wang. The global solution of the  $n$ -body problem. *Celestial Mechanics* 50, 73–88 (1991).
30. F. Diacu. The solution of the  $n$ -body problem. *The Mathematical Intelligencer* 18, 66–70 (1996).
31. K. Svozil. Omega and the time evolution of the  $n$ -body problem. *In*: Cristian S. Calude (ed.), *Randomness and Complexity, from Leibniz to Chaitin*. pp. 231–236. World Scientific, Singapore (2007) arXiv:physics/0703031.
32. F. Nietzsche. *Jenseits von Gut und Böse (Beyond Good and Evil)* (1886, 2009) digital critical edition of the complete works and letters, based on the critical text by G. Colli and M. Montinari, edited by Paolo D’Iorio. Berlin/New York, de Gruyter (1967).
33. F. Nietzsche. *Zur Genealogie der Moral (On the Genealogy of Morality)* (1887, 2009) digital critical edition of the complete works and letters, based on the critical text by G. Colli and M. Montinari, edited by Paolo D’Iorio. Berlin/New York, de Gruyter (1967).
34. G. Berkeley. *A Treatise Concerning the Principles of Human Knowledge*. (1710).
35. W.T. Stace. The refutation of realism. *In*: Herbert Feigl and Wilfrid Sellars (eds), *Readings in Philosophical Analysis*. pp. 364–372. Appleton-Century-Crofts, New York, 1949, previously published in *Mind* 53, 349–353 (1934).
36. T. Goldschmidt and K.L. Pearce. *Idealism: New Essays in Meta-physics*. Oxford University Press, Oxford, UK (2017, 2018).
37. C.S. Calude and G. Longo. The deluge of spurious correlations in big data. *Foundations of Science* 1–18, CDMTCS-488 (2016).

38. C.S. Calude and K. Svozil. Spurious, emergent laws in number worlds. *Philosophies* 4, 17 (2019), arXiv:1812.04416.
39. T.S. Kuhn. *The Structure of Scientific Revolutions*, 2nd ed. Princeton University Press, Princeton, NJ (1970).
40. P.K. Feyerabend. *Against Method*. New Left Books, London (1974).
41. P. Veyne. *Did the Greeks Believe in Their Myths? An Essay on the Constitutive Imagination*. University of Chicago Press, Chicago (1988).
42. C.A. Fuchs and A. Peres. Quantum theory needs no 'interpretation'. *Physics Today* 53, 70–71 (2000), further discussions of and reactions to the article can be found in the September issue of *Physics Today*, 53, 11–14 (2000).
43. R.P. Feynman. *The Character of Physical Law*. MIT Press, Cambridge, MA (1965).
44. B.-G. Englert. On quantum theory. *The European Physical Journal D* 67, 1–16 (2013), arXiv:1308.5290.
45. D.H. Wolpert. The lack of a priori distinctions between learning algorithms. *Neural Computation* 8, 1341–1390 (1996).
46. D.H. Wolpert and W.G. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* 1, 67–82 (1997).