
Incomputability at the foundations of physics (A study in the philosophy of science)

JOSÉ FÉLIX COSTA, *Instituto Superior Técnico, Universidade Técnica de Lisboa, Portugal; Centro de Matemática e Aplicações Fundamentais do Complexo Interdisciplinar da Universidade de Lisboa, and Centro de Filosofia das Ciências da Universidade de Lisboa.*
E-mail: fgc@math.ist.utl.pt

Abstract

In this article, we show that once we assume that there is some uniformity in Nature, empirical laws and theories involving the concepts of Physics can be learned by the scientists. We develop a mathematical framework to explain learnability of empirical relations. Then we classify the degree of learnability according to the degree of predictability by means of timescales. We also use the same mathematical framework to make formal the claim that learnability evolves according not only with the number of observations done, but also with the precision of measurements.

Keywords: Algorithmic scientific inference, incomputability, kolmogorov complexity, philosophy of science, predictive scientists, varying laws of physics.

1 Introduction

1.1 Motivation

In the last decade Barry Cooper (e.g., in [14–16]) has been stressing that the Turing universe can model the inevitable incomputability of the physical world. However, the Turing universe does not establish a clear link between Computability and Physics. In a sense, the recursion theoretic model of learning (that has something to say about deduction, induction or abduction from experiments) emphasizes some connections with Physics, although their mentors do not accept explicitly a possible non-computable reality (as stated in [8, 9]). In this work, we will try to unite these views of the world, introducing the idea that the ‘laws of nature’ are a computable manifestation of the incomputable. We will also introduce a degree of computability into the nature of the empirical physical law.

Algorithmic learning theory can be introduced just by looking at a quiz like this: $9 \rightarrow 72$, $5 \rightarrow 20$, $7 \rightarrow 42$, $8 \rightarrow 20$, $6 \rightarrow 30$, $3 \rightarrow ?$. To complete this sequence, to extrapolate to all positive integers, we attempt to find an algorithm and postulate a ‘mathematical law’ from a set of admissible ‘mathematical laws’. A similar activity is done by our scientist ‘Boyle’ when in a sequence of lab experiments on gases, while keeping the temperature constant, he measures the pressure of a gas against the volume it occupies.

The reader advances that the answer to the quiz is $3 \rightarrow 6$ and ‘Boyle’ succeeded in his paper on the physical law *pressure* \times *volume* = *some specific constant of the gas being used*. The quiz can be imagined as a sample of points of the graph of the function $n \rightarrow n \times (n - 1)$. The ‘Boyle’ quiz is a sample of the graph of the function *volume* \rightarrow *constant* / *volume*, for some value of the constant. An empiricist is able to identify a set of such laws from an infinite number of possible mathematical relations. Moreover, he identifies the law after having observed a finite sample of the graph of the function. Conventional scientists write mathematical formulas on paper using some

2 Incomputability at the foundations of physics

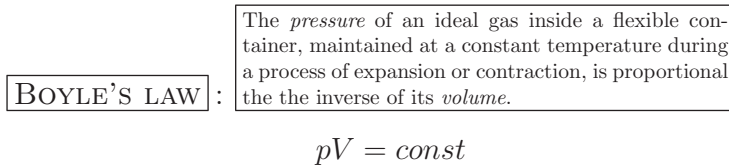


FIGURE 1. Scientist Boyle.

THE SCIENTIST OUTPUTS THE CODE e OF FUNCTION
--

```
Function pressure(volume :  $\mathbb{Q}$ ) :  $\mathbb{Q}$ ;  
Const  $K$  :  $\mathbb{Q}$ ; Constant of gas at some temperature  
Begin  
    Return  $K/\textit{volume}$   
End
```

FIGURE 2. The scientist writes the programme and outputs its code e . Notice that the rationals are encoded into the natural numbers.

universal grammar of symbolic expressions. In this article, the scientist writes an equivalent computer programme, outputting the code number of that programme (see Figure 1).

We will say that a ‘computer programme’ \mathcal{M} denotes the activity of a physicist if it explains some relations between concepts of Physics (such like the relation between *pressure* and *volume*). For example, \mathcal{M} , on inputting the observations of a dependent physical concept from an experiment (such like the looking-up table of *volume* \times *pressure* of a contained gas), outputs a new ‘computer programme’ $\{e\}$ that simulates and extrapolates the observations of the experience: the new ‘programme’, on input of the values assigned to the magnitudes of the independent physical concepts involved (such like the *volume*), outputs the predicted value of the dependent concept (such like the *pressure*). The right output is expected just after the input of finite data from the potential infinite information contained in the graph of the relation. For example, the ‘programme Boyle’, denoting a very restricted theory of physics (the theory of ideal gases undergoing isothermal processes), referred to in Figures 2 and 3, inputs the information about the observations of the pressure and volume of a contained gas, and outputs Boyle’s law for this gas. Such a law is a constitutive (empirical) relation between thermodynamical quantities which can be related to other thermodynamical quantities through a theory of thermodynamics. The output programme code e is the law of Physics instantiated to that particular gas. We expect that the scientist can reach code e after finitely many observations.

The same ‘programme’ may identify a countable number of physical laws, e.g. obtained by closure under derivation. A ‘programme’ may represent a physical theory. In a strict sense, a ‘programme’ may explain *an infinite set of instances of a physical law*, such as ‘Boyle’ does (see Figure 3). We refer to the ‘programme’ as a *scientist*. Thus, we say that a scientist *identifies* a physical law if he/she identifies the instances of the physical law.

In general, a scientist can identify *different instances of the same or of different physical laws*.

Ultimately, a physical theory is represented by a scientist capable of identifying the observable consequences of the theory that correspond to a set of functions expressing non-causal relations or either weak or strong causal laws (not necessarily of the same type¹) about physical observables, such as the celestial positions of planets in the solar system as a function of time.

¹In the sense that the hyperbolic, parabolic and elliptic orbits of the celestial bodies belong to different curve types, although, they belong to all the same family of conic curves. It follows that the scientist who identifies elliptic orbits of planets

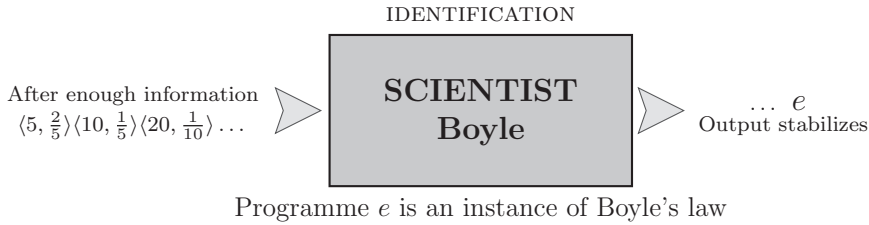


FIGURE 3. For all $t \geq p$, scientist \mathcal{M} on input $Volume[Pressure]$ outputs the instance of Boyle's law for the particular ideal gas under consideration: $pressure \times volume = \text{some specific constant}$. This sampling process with its own computability theoretic character illustrates the empiricist according to Definition 2.1.

In this article, we discuss a model of learning the laws of Physics, having the following ideas in mind:

- (1) The (empirical) scientist learns a relation between concepts of physics from a finite sample of observations and measurements.²
- (2) The scientist learns a theory if he/she is able to identify a (possibly infinite) set of relations between concepts pertaining to that theory. Some relations may be deduced and used as refutation tools or basic statements (in the sense of Popper); some others are extrapolated from observations and controlled experiments.
- (3) A law of physics is a universal relation between concepts, possibly involving universal parameters, such like the universal constant of ideal gases, or the velocity of light.
- (4) Although 'universal' an empirical law of physics might evolve according with the precision of the measurements (e.g., predicting the orbit of Mercury from Newtonian theory of gravitation to Einstein field equations; e.g., a 'universal constant' becomes a slow varying parameter).
- (5) The law is ultimately denoted by a computer programme that simulates the relation between the concepts and forecasts the physical system.

Important remark: in this article, the use of Physics could be replaced by any other empirical science equipped with a measurement system a la Hempel (see [21]). However, we prefer to provide examples from just one field.

1.2 Notation

In what follows, by FUN we denote the set of all total functions (computable and non-computable) of type $\mathbb{N} \rightarrow \mathbb{N}$. By \mathcal{R} we denote the set of total computable functions, by \mathcal{R}_0 we denote the set of computable functions generated by loops type FOR (and never loops type WHILE), i.e., the set of primitive recursive functions. In general, we talk about sets of functions of type $\mathbb{N} \rightarrow \mathbb{N}$. For example, the set \mathcal{AEZ} of unary computable functions which take the value zero almost everywhere (Almost Everywhere Zero) is an abstraction in the limit of some meaningful law of physics, described by relations between two arbitrary variables that satisfy such a property, which includes the relations expressing the characteristics of the sets of critical temperatures at which substances

may not identify the 'parabolic' orbits of comets, i.e, the scientist that distinguishes planets from comets and asteroids is less trivial than a scientist that identifies only planets. In this matters, we fall into the discussion of Popper in [26], Section 36.

²In Section 7, a note is added on theorists.

4 Incomputability at the foundations of physics

suffer a (generalized) phase transition. Such an example is not less abstract than saying that the set of functions $y=f(x)$, such that $xy=constant$, is the set of instances of Boyle's law of perfect gases at constant temperature ($pressure \times volume = constant$), and not less abstract than saying that the set of functions $yx^2 = constant$ encode Newtonian fields. *Particular examples from physics* should not appear to constitute an obstacle to the mathematics developed in a more abstract level. For example, the set \mathcal{SD} of all (total) computable functions ψ such that $\psi(0)$ is the code of a programme for ψ (Self Describing) can be seen as the set of laws of physics (as number codes) followed by their own graphs. By ϕ_e we denote the e -th (partial) computable function, according with any suitable encoding into the natural numbers.

1.3 Claims

In this article, we prove that there exists an infinite non-collapsing hierarchy of classes $PRED_\alpha$ of predictive theories of different ordinal degree α of incomputability, i.e., classes of sets of computable and non-computable relations containing the known classes EX and BC (to be surveyed in Section 2.1) of sets of learnable computable relations.

PROPOSITION 1.1

For all ordinal $\alpha \in \omega^\omega$,

$$EX \subset BC \subset \dots \subset PRED_{\alpha+1} \subset PRED_\alpha \subset \dots \subset PRED_0$$

Moreover, all the new classes contain the set of computable functions.

PROPOSITION 1.2

$\mathcal{R} \in PRED_\alpha$, for all ordinal $\alpha \in \omega^\omega$.

Even before introducing the concepts we informally advance the statement.

PROPOSITION 1.3

For every ordinal $\alpha \in \omega^\omega$, 'a law' in $PRED_\alpha$ departs from some 'law' in \mathcal{R} only after time greater than $2^\alpha(t)$, where t is the size of the observations so far done and $2^\alpha(t)$, for $\alpha \geq \omega$, is a function that grows fast enough.

The structure of the article is the following: in Section 2, we survey very briefly how learning theory can be applied to understand empirical science; in Section 3, we develop the basics of timescales to model the range of application of physical theories; in Section 4, we generalize the established EX and BC frameworks to deal with non-computable phenomena; the characterization of the new predictive scientists is done in Sections 5 and 6; finally, in Section 7, we add some thoughts as concluding remarks.

2 Empirical science

2.1 Data

Empirical observations appear as dyadic rational numbers of at most ℓ decimal places long (see Figure 4). The number of decimal places denote the precision of the observations. The set of all observations can be taken as one to one encoding of the rational numbers of ℓ decimal places into the natural numbers. However, this *DATA* may not be provided to the scientist in ascending order

CELESTIAL LONGITUDE : Pattern of observations of positions of the planets in the sky in radians as a function of time up to ℓ bits of precision:

$$\psi(\underbrace{\text{encoding-of-}(t_0, t_1 \dots t_\ell)}_\ell) = \text{encoding-of-}(\underbrace{\tilde{\psi}(t)_0, \tilde{\psi}(t)_1 \dots \tilde{\psi}(t)_\ell}_\ell)$$

FIGURE 4. The pattern of a variable or of a parameter.

of some independent variable. We will discuss the relevance of the ordering of observations later on in Sections 2.2 and 4. For example, the celestial longitude of a planet is provided by the angle $\tilde{\psi} : \mathbb{Q} \rightarrow \mathbb{Q}$, as a function of time t with a fixed precision of ℓ decimal places, encoded into a function $\psi : \mathbb{N} \rightarrow \mathbb{N}$.

Information about empirical observations appear in general texts T , such as look-up tables.

A text T for a function ψ is a map of type $\mathbb{N} \rightarrow [(\mathbb{N} \times \mathbb{N}) \cup \{\#\}]$, where the elements of the graph of ψ , $\langle t, \psi(t) \rangle$, for $t, \psi \in \mathbb{N}$, are given separated by #. We assume that each pair is such that no other pair $\langle t, \psi'(t) \rangle$ exists in the text with $\psi(t) \neq \psi'(t)$. Repetitions are allowed. Order of elements can be, in principle, arbitrary. We count the elements from a text starting with 0 and we denote by $T(t)$ the t th element of the text, i.e., the element number $t - 1$, where separators do not count in the counting. By $T[t]$ we denote the sequence of the first t elements of T . Sometimes we abbreviate $T[t]$, where T is the text for the function ψ , by $\psi(0)\#\psi(1)\#\dots\#\psi(t-1)\#$, instead of the full writing $\langle 0, \psi(0) \rangle \# \langle 1, \psi(1) \rangle \# \dots \# \langle t-1, \psi(t-1) \rangle \#$, assuming that the pairs of the graph of ψ are given in order (canonical presentation). Variables for prefixes of text T are generally σ, τ, ρ , etc. $DATA = \{T[t] : T \text{ is a text for a function and } t \in \mathbb{N}\}$ is the set of all prefixes of text for functions. Concatenation of two sequences $\sigma, \tau \in DATA$ is represented by $\sigma \diamond \tau$.³ Using more or less standard notation, we write $\sigma \subset \tau$ to say that σ is a proper prefix of τ and $\sigma \subseteq \tau$ to say that σ can coincide with τ .⁴ By $content(\sigma)$ we denote the set of pairs occurring in σ , if $\sigma \in DATA$. We can also refer to the finite function given by $\sigma \in DATA$ as $\hat{\sigma}$, and to its domain as $dom(\hat{\sigma})$. The range of a function $f : \mathbb{N} \rightarrow \mathbb{N}$ is denoted by $\rho(f)$. We also refer to the number of pairs in σ , as the size of σ , i.e., $|\sigma|$.

Notice that we denote by ϕ_e the (partial) computable function computed by the programme of code e , assuming a standard programming system (such like that of Turing machines) encoded into the natural numbers as usual.

2.2 Scientists and scientific inference

An attempt to a general definition of a *scientist* appeared in [20] as follows.

DEFINITION 2.1 (Scientist, Gold [20])

A scientist is a function (possibly partial, not necessarily computable) of type $DATA \rightarrow \mathbb{N}$.

The scientist take data from $DATA$ and produce programme codes; by some kind of inference⁵ from a finite number of observations, each scientist outputs natural numbers to be interpreted as programme codes that compute ‘empirical laws’ (computable functions) of type $\mathbb{N} \rightarrow \mathbb{N}$. A scientist will converge into an instance of a physical law (see Figure 5), if that physical law turns to predict, i.e. to extrapolate, the future behaviour of the physical system, e.g. to compute the trajectory of each

³Note that the concatenation of two sequences of $DATA$ may not belong to $DATA$.

⁴In these cases, the symbol # does not count, i.e., if $\sigma \subset \tau$ then σ contains less pairs of numbers than τ .

⁵For example, either because the scientist holds a theory of physics and makes a deduction, or because he makes the inference by any other different method.

6 Incomputability at the foundations of physics

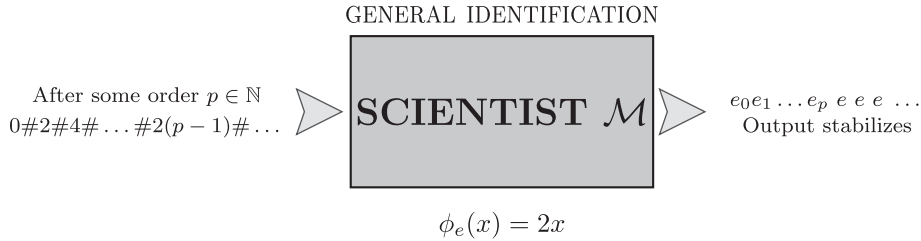


FIGURE 5. For all $t \geq p$, scientist \mathcal{M} on input $\psi[t]$ outputs a code e of ψ .

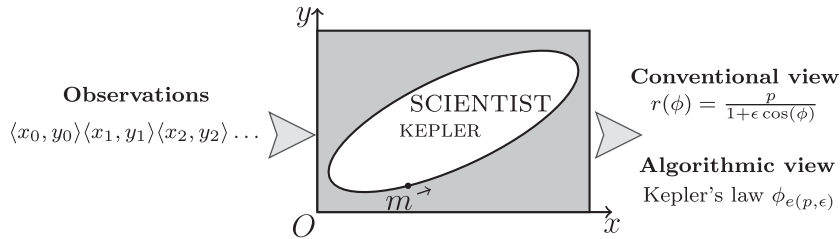


FIGURE 6. Kepler's first law. The the ellipse is provided indirectly by a code $\phi_{e(p, \epsilon)}$ as a function within, let us say, the collection of all such ellipses or the collection of all elementary functions of Analysis.

planet from an observed portion of it. For example, a scientist, holding the Newtonian theory of Gravitation, can predict an orbit if enough information is provided as input. Or, otherwise, he can apply Gaussian methods to predict an orbit from a class of possible orbits.

One conspicuous event in the history of planetary science is the discovery of planet Neptune. After Herschel's discovery of Uranus, the deviations of Uranus from the computed orbit led to successive recalculations of its orbit. But the planet always failed to meet the computed orbit. Uranus escaped computation since some further information was needed such as the existence of a new planet Neptune.

DEFINITION 2.2 (Scientific success on a single function)

Let $\psi : \mathbb{N} \rightarrow \mathbb{N}$ be a total function. We say that the scientist \mathcal{M} *EX-identifies* ψ if there exists an $e \in \mathbb{N}$ and an order p such that, for $t \geq p$, $\mathcal{M}(\psi[t]) = e$ and $\phi_e = \psi$.

Returning to Figure 5, we can see that a scientist can change his/her mind and produce different hypotheses e_0, e_1, \dots, e_p, e , for different input prefixes of the same text. What is required in *EX-identification* is that after acquiring sufficient knowledge about the phenomenon, the scientist converge into a single stable hypothesis e .

Figure 6 illustrates the fact that the output of a scientist is not a law in analytic terms (the equation of an ellipse) but the programme code $\phi_{e(p, \epsilon)}$ of such a law.

We can propose an alternative definition of the concept of identification.

DEFINITION 2.3 (Scientific success on a single function)

Let $\psi : \mathbb{N} \rightarrow \mathbb{N}$ be a total function. We say that scientist \mathcal{M} identifies ψ if there exists an $e \in \mathbb{N}$ and numbers $p, \ell \in \mathbb{N}$ such that, for $t \geq p$, $\mathcal{M}(\psi[t]) = e$ and, for all $t \in \mathbb{N}$, $|\phi_e(t) - \psi(t)| \leq 2^{-\ell}$.⁶

⁶In the standard context of learning theory, just take $\ell = +\infty$, and we have \mathcal{M} converging to e on $\psi[t]$ and, for all $t \in \mathbb{N}$, $\phi_e = \psi$.

Note that, not considering the case of infinite precision, we just want to guaranty that the predicted trajectory falls, e.g. in the limit of the naked eye observations.⁷ In this sense, on precision, Karl Popper says:

Assume that the consequences of two theories differ so little in all fields of application that the very small differences between the calculated and observable events cannot be detected, owing to the fact that the degree of precision attainable in our measurements is not sufficiently high. It will then be impossible to decide by experiments between the two theories, without first improving our technique of measurements. This shows that the prevailing technique of measurement determines a certain range — a region within which discrepancies between observations are permitted by the theory.

As an example of *EX*-identification, and to introduce also an example of a learning algorithm, consider the set \mathcal{AEZ} of (total) computable functions ψ identical to 0 almost everywhere. Just consider the scientist \mathcal{M} that, on input $\psi(0)\#\dots\#\psi(t-1)\#$, builds the canonical (ordered, finite) list μ of the pairs $\langle t, \psi(t) \rangle$, such that $\psi(t)$ is non-zero, and outputs *the code of the programme*

$$\lambda x. \text{If } x \in \text{dom}(\widehat{\mu}) \text{ Then } \widehat{\mu}(x) \text{ Else } 0 .$$

Such a scientist *EX*-identifies all functions of \mathcal{AEZ} according to Definition 2.3.

In yet another identification paradigm, the scientist can keep changing his/her mind about the phenomenon being analysed, by outputting different codes. However it is required that, after some order, after acquiring sufficient knowledge about the function being modeled, the scientist only produces correct codes. Although it might not seem, these scientists are more powerful than *EX*-scientists.

DEFINITION 2.4 (*BC*-identification, Bärzdiņš [4], Case and Smith [13])

A scientist \mathcal{M} *BC*-identifies $\psi \in \mathcal{R}$ if, for all but finitely many n , we have $\phi_{\mathcal{M}(\psi[n])} = \psi$.

The letters in *BC* stands for **B**ehaviourally **C**orrect.

For a collection of total functions (e.g. a collection of planetary orbits), possibly infinite, we have in any case of identification.

DEFINITION 2.5 (Scientific success on a collection of functions)

We say that scientist \mathcal{M} *EX*-, *BC*-identifies a set S of total functions just in case he/she *EX*-, *BC*-identifies every $\psi \in S$. By *EX* and *BC* we denote the class of sets *EX*-, *BC*-identifiable, respectively.

PROPOSITION 2.6 (Blum and Blum [6], Case and Smith [12, 13], Jain et al. [22])

The class *EX* is not closed under union.

For example, a function can belong to \mathcal{AEZ} and not belong to \mathcal{SD} ; a function can belong to \mathcal{SD} and not belong to \mathcal{AEZ} ; but there are functions which share the property of \mathcal{AEZ} and \mathcal{SD} at the same time. The set $\mathcal{SD} \cup \mathcal{AEZ}$ is an example of a set not belonging to *EX*, although both \mathcal{SD} and \mathcal{AEZ} are in *EX*.

As consequence, the set of all (total) computable functions \mathcal{R} is not *EX*-identifiable, because if it were, then *EX* would be closed for union.

⁷We are simplifying notation: $|\phi_e(t) - \psi(t)|$ means the difference between the computed and the observed trajectories in terms of *rational numbers*; thus $\psi(t)$ means both the rational coordinate and its code; whenever rational values appear, such as $2^{-\ell}$, all the quantities involved should be interpreted as rational numbers.

8 Incomputability at the foundations of physics

PROPOSITION 2.7

$\mathcal{R} \notin EX$.

PROPOSITION 2.8 (The hierarchy of scientists, mainly from Case and Smith [12, 13])

$$\mathcal{R} \notin EX \subset BC \not\equiv \mathcal{R}$$

The idea is that sets A and B may be EX -identified without having $A \cup B$ EX -identified, meaning that two different algorithmic theories may not have an unification.

Proposition 2.6 tells us that unification is not always possible.

THEOREM 2.9 (Unification of scientific laws)

Unification of scientific laws (as algorithms) is not always possible within the paradigms EX or BC .

2.3 Scientists: total and Popperian

Let \mathcal{M} be a scientist aiming at explaining the motions of the planets as seen projected in the celestial sphere. On input $\psi[t] = \psi(0)\# \dots \#\psi(t-1)\#$, \mathcal{M} outputs the programme $e_{\psi[t]}$. For some value of t , $e_{\psi[t]}$ becomes constant, that is $e_{\psi[t]} \rightarrow e$. The index e codes for the programme that extrapolates the trajectory of a given planet as a function of time. The Popperian attitude would be to search for a refutation of the physical law encapsulated in the scientist \mathcal{M} . Finding a lack of agreement in the predicted trajectory is an error of commission according to Daley ([17, 22]; not providing any answer at all (in finite time) for the position of the planet at instant t is an error of omission according to Daley ([17, 22])). Possible errors of omission were not, however, considered by Popper, resulting in a possible definition of *Popperian scientist*.

DEFINITION 2.10 (Popperian, Case and Ngo-Manguelle [10, 11])

We say that a scientist \mathcal{M} is *Popperian* on (total) data $f: \mathbb{N} \rightarrow \mathbb{N}$ if, for all $t \in \mathbb{N}$, $\phi_{\mathcal{M}(\psi[t])} \in \mathcal{R}$. We say that a scientist is Popperian on a set S if \mathcal{M} is Popperian on each function of S . Finally, \mathcal{M} is Popperian if \mathcal{M} is Popperian on the whole set of computable functions \mathcal{R} .

We provide a proof of two very basic results, stated here and there, about scientists that help to understand the concepts so far introduced and some basic proof methods to reason about scientists.

PROPOSITION 2.11

If $\mathcal{M}^0, \mathcal{M}^1, \dots, \mathcal{M}^i, \dots$ is an enumeration of all computable scientists, then there exists a (total) computable function $f: \mathbb{N} \rightarrow \mathbb{N}$ such that, for all $i \in \mathbb{N}$: (a) $\mathcal{M}^{f(i)}$ is a total scientist and (b) if \mathcal{M}^i identifies the (total) computable function ψ , then $\mathcal{M}^{f(i)}$ also identifies ψ .

PROOF. We consider any (either total or partial) of the scientists \mathcal{M}^i , assuming that he identifies a given function ψ , and we algorithmically specify a new total scientist $\mathcal{M}^{f(i)}$ that identifies ψ .

Let T be any text for ψ . On input $T[t]$, the new scientist $\mathcal{M}^{f(i)}$ calls the scientist \mathcal{M}^i to run for t steps, successively and orderly, on all the prefixes of $T[t]$. Then the scientist $\mathcal{M}^{f(i)}$ takes the longest prefix $\tilde{\sigma}$ of $T[t]$ for which \mathcal{M}^i outputted a hypothesis in t steps. In case of no hypothesis have been produced in t steps for any prefix, the new scientist $\mathcal{M}^{f(i)}$ outputs 0, otherwise he outputs $\mathcal{M}^i(\tilde{\sigma})$.

Scientist $\mathcal{M}^{f(i)}$ is total and computable. The conversion algorithm is synthesised in Figure 7.

Let us now suppose that T is a text for a function ψ . Then, there exists an order $p \in \mathbb{N}$ such that scientist \mathcal{M}^i , on input $T[p]$, converges to the final hypothesis and there exists an order $q \geq p$ such that, for $j \geq q$, j steps are enough for \mathcal{M}^i to converge on $T[p]$. Thus, after order q , $\mathcal{M}^{f(i)}$ outputs the final hypothesis on all prefixes of text T of size greater or equal to q . ■

$\mathcal{M}^{f(i)}$: Function $\mathcal{M}^{f(i)}(\sigma : SEQ) : \mathbb{N}$; Begin $t := \sigma $; $j := 0$; For $k := 1$ To t Do If $\mathcal{M}^i(\sigma[k]) \downarrow \leq t$ Then $j := k$; If $j = 0$ Then Output 0 Else Output $\mathcal{M}^i(\sigma[j])$ End
--

FIGURE 7. The total scientist $\mathcal{M}^{f(i)}$ constructed from the possibly partial scientist \mathcal{M}^i .

PROPOSITION 2.12

If a scientist \mathcal{M} converges to e on the canonical text for the function $\psi = \phi_e$, then there exists a scientist $\tilde{\mathcal{M}}$ that converges to e on all texts for ψ .

PROOF. This result is a consequence of the fact that only the total functions are considered for identification.

Let T be an arbitrary text for ψ and $\tilde{\mathcal{M}}$ be the scientist that, on input $\sigma \subset T$, computes first the longest prefix $\tilde{\sigma} \in DATA$ such that $dom(\tilde{\sigma}) \subseteq dom(\sigma)$ and then calls \mathcal{M} on $\tilde{\sigma}$. Note that such prefixes can always be formed even if they require very long input sequences σ . Note also that, as $|\sigma| \rightarrow \infty$, we have both $\sigma \rightarrow T$ and $\tilde{\sigma} \rightarrow T$. Then, we conclude that, if \mathcal{M} converges to e on the canonical text for ψ , then scientist $\tilde{\mathcal{M}}$ converges to e on all texts for ψ . ■

In the general case we will be considering in Section 4, the order of the empirical observations, i.e., the ordering of the elements in the text, will be considered (possibly) relevant since it may take into account the dependence between different such observations, such like timed relations.

3 Timescales

3.1 Families of timescales

In this section, we introduce the concept of a timescale.

By $(\cdot)_{10} : \{0, 1\}^* \rightarrow \mathbb{N}$ we designate the operation of reading the string enclosed in brackets in decimal⁸ and by $(\cdot)_2 : \mathbb{N} \rightarrow \{0, 1\}^*$ we designate the operation of reading the natural number enclosed in brackets in binary.

A bijection $\iota : \{0, 1\}^* \rightarrow \mathbb{N}$ can easily be defined as $\iota(w) = (1w)_{10} - 1$.

Let $(\cdot)^- : \{0, 1\}^* \rightarrow \{0, 1\}^*$ denote the partial function that removes the leftmost bit from a given non-empty binary sequence w . Thus, the inverse function of ι is $\kappa : \mathbb{N} \rightarrow \{0, 1\}^*$ such that $\kappa(t) = [(t+1)_2]^-$.⁹

DEFINITION 3.1 (Timescale)

We say that a (total) computable function $\pi : \mathbb{N} \rightarrow \mathbb{N}$ is a **timescale** if π is a non-decreasing surjective (linear or) sublinear function.

⁸For example, 00100 reads 4 in decimal.

⁹For example, the string 001 is first mapped into 1001 and then into 8. In the other way round, the number 9 is first mapped into $8+1=9$, then into 1001, and finally into 001, by removing the leftmost 1.

10 Incomputability at the foundations of physics

Remember that we write $\psi \in o(\varphi)$ whenever for all $\eta \in \mathbb{R}^+$, there exists an order $p \in \mathbb{N}$ such that, for all $t \geq p$, $\psi(t) < \eta\varphi(t)$. Another (equivalent) way of characterizing $o(\varphi)$ is by limit taking:

$$o(\varphi: \mathbb{N} \rightarrow \mathbb{N}) = \left\{ \psi: \mathbb{N} \rightarrow \mathbb{N}: \lim_{t \rightarrow \infty} \frac{\psi(t)}{\varphi(t)} = 0 \right\}$$

Remember also that we write $\psi \in O(\varphi)$ whenever there exists $\eta \in \mathbb{R}^+$ and an order $p \in \mathbb{N}$ such that, for all $t \geq p$, $\psi(t) < \eta\varphi(t)$. We define $\psi \in O_s(\varphi)$ as the set $\{\psi \in O(\varphi): \psi \text{ is a timescale}\}$

DEFINITION 3.2 (Ordering timescales)

There is a relation between two total functions, $\psi, \varphi: \mathbb{N} \rightarrow \mathbb{N}$, by saying that $\psi \prec \varphi$ if $\psi \in o(\varphi)$. This relation can be generalized to two classes of functions, \mathcal{F} and \mathcal{G} , by saying that $\mathcal{F} \prec \mathcal{G}$ if there exists a function $\varphi \in \mathcal{G}$, such that for all functions $\psi \in \mathcal{F}$, $\psi \prec \varphi$.

DEFINITION 3.3

A class of timescales \mathcal{F} is said to be *reasonable* if $\mathcal{F} = O_s(\mathcal{F})$.

Examples of timescales are the logarithm function and the successive iterations of the logarithm function. Define \log as the function such that on inputs greater or equal to 1 behaves exactly as the conventional \log function, but for inputs less than 1 outputs 0:

$$\log(t): \mathbb{R} \mapsto \mathbb{R} := \begin{cases} \log(t) & \text{if } t > 1 \\ 0 & \text{if } t \leq 1 \end{cases} \quad (\text{for } t \in \mathbb{R})$$

To work over the non-negative integers we redefine:

$$\log(t): \mathbb{N} \mapsto \mathbb{N} := \begin{cases} \lceil \log(t) \rceil & \text{if } t \geq 1 \\ 0 & \text{if } t = 0 \end{cases} \quad (\text{for } t \in \mathbb{N})$$

DEFINITION 3.4 (Ordinal iterate of the logarithm)

For each ordinal α , we define inductively the class $\log^{(\alpha)}$: (a) if α is 0, then $\log^{(0)} = O_s(\lambda t.t)$, (b) if α is a successor ordinal, then

$$\log^{(\alpha+1)} = O_s(\{\lambda t.\log(\psi(t)): \psi \in \log^{(\alpha)}\}),$$

and (c) if α is a limit ordinal, then

$$\log^{(\alpha)} = O_s(\cap_{\gamma \in \alpha} \log^{(\gamma)}).$$

Successive iterations of the logarithm function generate an infinite family of classes of reasonable timescales. Note that, trivially, we have.

PROPOSITION 3.5

For all $\alpha \in \omega$, $\log^{(\alpha+1)} \prec \log^{(\alpha)}$.

Then Proposition 3.5 allows us to take the infinite descending chain of timescales.

PROPOSITION 3.6 (A first hierarchy of timescales)

$\log^{(\omega)} \prec \dots \prec \log^{(3)} \prec \log^{(2)} \prec \log^{(1)} \prec \log^{(0)}$.

To see that $\log^{(\omega)}$ is not trivial, we add the following:

PROPOSITION 3.7 (Non-triviality of $\log^{(\omega)}$)
 $\log^{(\omega)} \neq \emptyset$.

PROOF. Consider the function \log^* , defined by (a) $\log^*(t)=0$, for $t=0$, and (b) $\log^*(t)=\min\{k : \log^{(k)}(t) \leq 1\}$, for $t > 0$. ■

To continue the descending chain we define the classes $\log^{(\omega+k)}$, for $k \geq 1$, to be the class generated by $\log^{(k)} \circ \log^{(\omega)}$. Again we take the limit $\log^{(2\omega)} = O(\bigcap_{k \geq 1} \log^{(\omega+k)})$.

PROPOSITION 3.8 (A second hierarchy of timescales)
 $\log^{(2\omega)} \prec \dots \prec \log^{(\omega+1)} \prec \log^{(\omega)} \prec \dots \prec \log^{(2)} \prec \log^{(1)} \prec \log^{(0)}$.

PROPOSITION 3.9 (Non-triviality of $\log^{(2\omega)}$)
 $\log^{(2\omega)} \neq \emptyset$.

PROOF. Consider $\log^{*(2)} = \log^* \circ \log^*$. ■

Now the function $\log^{*(2)} = \log^* \circ \log^*$ is in $\log^{(2\omega)}$, so the class $\log^{(2\omega)}$ is not trivial. We can continue descending by setting $\log^{(2\omega+k)}$, for $k \geq 1$, to be the class generated by $\log^{(k)} \circ \log^{*(2)}$.¹⁰ Of course this continues until we reach

$$\log^{(\omega^2)} = O_s(\bigcap_{k \geq 1} \log^{(k\omega)}).$$

To get beyond this would require finding a function in $\log^{(\omega^2)}$. But, for the purpose, we only have to diagonalize over $\log^{*(1)}$, $\log^{*(2)}$, $\log^{*(3)}$, ..., to obtain \log^{**} , the same way we obtained \log^* :

$$\log^{**}(t) = \min\{k : \log^{*(k)}(t) \leq 1\},$$

with $\log^{**}(0)=0$. Such a reasoning can be repeated and the chain of classes eventually reduce to the empty set at ω^ω . However, *we are only concerned with the non-collapsing character of the chain until some (transfinite) ordinal*.

To finish the section, we add a few more notation. To each timescale function π we associate one right inverse π^{-1} , i.e. such that $\pi \circ \pi^{-1}(t)=t$. For example, if $\pi(t) = \lceil \log(t) \rceil$, then we take $\pi^{-1}(t) = 2^t$. Note also that, in general, the left inverse does not exist, since π functions have plateaux, i.e., they are non-decreasing, but not strictly monotonic, and, consequently, not injective.

3.2 A note on Kolmogorov complexity of infinite sequences

Kolmogorov complexity is not needed in the study that follows. However, since the basic concepts of advice class and Kolmogorov class are related, we provide, in this short subsection, a discussion of a few concepts namely the definition of *Kolmogorov complexity* given in [1]. We will see in Section 6 how Kolmogorov complexity fits to learning theory.

DEFINITION 3.10 (Kolmogorov complexity of a stream, Balcázar and Hermo [3])

Let \mathcal{U} be a universal Turing machine, let $\pi : \mathbb{N} \rightarrow \mathbb{N}$ be a total function and $g : \mathbb{N} \rightarrow \mathbb{N}$ be a time constructible function, and let $\alpha \in \{0, 1\}^\omega$. We say that α has Kolmogorov complexity $\mathcal{K}[\pi, g]$ if there exists $\beta \in \{0, 1\}^\omega$ such that, for all n , the universal machine \mathcal{U} outputs $\alpha[n]$ in time $g(n)$, when given n and $\beta[\pi(n)]$ as inputs (see Figure 8).

¹⁰By $\log^{(k)}$ we denote both, the iterated logarithm and the class of functions according to Definition 3.5.



FIGURE 8. Predicting n bits of α being helped by $\pi(n)$ bits of β .

This definition can be restated as follows: the binary sequence $\alpha[t]$ of size t is generated by a universal Turing machine given the binary sequence $\beta[\pi(t)]$ as input. The reader should look to the input $\beta[\pi(t)]$ as a binary sequence made of a shorter prefix, γ , which is the required programme to the universal Turing machine, paired with the actual input w ($\beta[\pi(t)] = \gamma w$). In what follows, the required programme coded in γ turns to be the code of the scientist himself given separately, being the sequence β pure advice $\beta: \mathbb{N} \rightarrow \{0, 1\}$. $K[\pi, g]$ can also be seen as the set of all infinite binary sequences with Kolmogorov complexity $K[\pi, g]$. By $K[\pi]$ we denote the set of all infinite binary sequences with Kolmogorov complexity $K[\pi, g]$, where g is an arbitrary time constructible function. $K[\mathcal{F}]$ is the set of all infinite binary sequences taken from $K[\pi]$, where $\pi \in \mathcal{F}$.

4 The concept of algorithmic scientific theory

[Changing laws of Physics, Peirce [25]] Now the only possible way of accounting for the laws of Nature and for uniformity in general is to suppose them result of evolution. This supposes them not to be absolute, not to be obeyed precisely. It makes an element of indeterminacy, spontaneity, or absolute chance in nature.

To model the predictive but limited power of scientific theories, we will introduce new classes that generalise the **EX**planatory class of Gold ([20]), investigated by Case and Smith (in [12, 13], and with more modern notation by Jain et al. in [22]), and the **Behaviourally Correct** class of Bärzdiņš ([4]), investigated by Feldman ([18]) and Case and Smith ([13]).

We assume that a law of physics corresponds to the extraction of algorithmic content from (possible non-computable) phenomena to predict Nature to some extent. Thus, our scientists have to be equipped with some timescale π (being it large, medium or small) such that, prediction of observations at values $0, 1, \dots, t-1$, may require data at values $0, 1, \dots, \pi(t)-1$, e.g., to predict the values of some function until $t-1$ (t values), the scientist requires advice data of size at most $\log(t)$. We might admit that the text is given in the order of the natural numbers (cf. Proposition 2.12).

DEFINITION 4.1

The data (text) for function ψ up to time t is the sequence $\psi[t] = \psi(0)\#\psi(1)\#\dots\#\psi(t-1)\#$, where $\psi(i)$ expresses the observations at time i .

Note that, when we talk about the text for the Boyle's law, such as in Figure 3, we are assuming that the text comes in arbitrary order, as arbitrary as the order of different experiments involving the measurement of *volume* and *pressure*. We are, indeed, assuming the independence of different observations. If we want to include the dependence on the order of experiments, we would be led to inputs of the form $p[V] = \langle V_0, p_0 \rangle \# \dots \# \langle V_{t-1}, p_{t-1} \rangle \#$, for $t = \text{time}$, $V = \text{volume}$, and $p = \text{pressure}$. In such texts, the order of the pairs *volume/pressure* would be irrelevant! However, order might become fundamental, e.g., in the context of non-repeatable conditions, such as the measurement of a celestial longitude, or other phenomena involving time.

The most general model we propose is based on an idea of a compromise between the conventionalist approach to science (represented by Poincaré and Pierre Duhem), Popper’s view of refutability and Peirce calling for the evolving laws of Physics. In the words of Popper ([26]):

According to this conventionalist point of view, laws of nature are not falsifiable by observation; for they are needed to determine what an observation and, more especially, what a scientific measurement is. It is these laws, laid down, by us, which form the indispensable basis for the regulation of our clocks and the correction of our so-called ‘rigid’ measuring-rods. A clock called ‘accurate’ and a measuring rod ‘rigid’ only if the movements measured with the help of these instruments satisfy the axioms of mechanics which we have decided to adopt.

This means, *after the event*, that the necessary inaccuracy of measurements can hide some non-computability even in the laws of macrophysics, in a way such that more accuracy translates into revision of scientific paradigms. This is what happened with Tycho Brahe’s naked eye observations in the XVI century, which caused a truly revision of the celestial machinery. This is also what was attempted at the time of the discovery of planet Neptune in the XIX century, which was about to cause a change in the theory of gravitation. And what happened, after all, with the advance of Mercury’s perihelion by the end of the XIX century. The non-computability that might be conceived to be hidden in inaccuracy of measurements (implying too that a large departure from regularity would have constitute an obstacle to the scientific achievement), can be modelled by a timescale of application of the physical law.

Let β be any ω -sequence of type $\mathbb{N} \rightarrow \{0, 1\}$. Let \mathcal{R}^β denote the least set of functions ψ containing the 0-ary zero Z , the unary successor function S , the projection functions, the given function β , and closed under composition, primitive recursion and minimalization, i.e., the set of computable functions relative to β . This set can be encoded into the natural numbers. By ϕ_e^β we denote the e th-function in that set. A relativised Kleene’s second recursion theorem applies.

DEFINITION 4.2

Let π be a timescale. The class $\mathcal{R}_\pi^\beta \subset \mathcal{R}^\beta$ contains all the total functions $\psi \in \mathcal{R}^\beta$ such that, for all $t \in \mathbb{N}$, the values of $\psi(0), \psi(1), \dots, \psi(t-1)$ only depend on the values of $\beta(0), \beta(1), \dots, \beta(\pi(t)-1)$.

Suppose that the function β is non-computable and we want to know the values of ϕ_e^β for $x = 0, 1, \dots, \tau - 1$. We can just redefine ϕ_e^β , providing the programme $\{e\}$ with the computable β' such that β' coincide with β for $x = 0, 1, \dots, \pi(\tau)$ and $\beta'(t) \neq \beta(\pi(\tau))$ for $t \geq \pi(\tau) - 1$. The new function $\phi_{e'}$ is such that $\phi_{e'}[\tau] = \phi_e^\beta[\tau]$. We say that programme $\{e'\}$ predicts ϕ_e^β for $t = 0, 1, \dots, \tau - 1$.

DEFINITION 4.3 (Predictive scientist, Figure 9)

Let $\pi : \mathbb{N} \rightarrow \mathbb{N}$ be a scale function and $\beta : \mathbb{N} \rightarrow \{0, 1\}$ be an advice sequence. We say that a scientist \mathcal{M} $PRED_\pi^\beta$ -identifies the (total, possibly non-computable) function ψ if there exists an order $p \in \mathbb{N}$ such that, for all $t \geq p$, \mathcal{M} on input $\psi[\pi(t)]$, and using at most the prefix $\beta[\pi(t)] \subset \beta$ of the advice, outputs a programme code for $\tilde{\psi}$ such that, for every $i \leq t$, we have $\psi(i) = \tilde{\psi}(i)$.

DEFINITION 4.4 (Success with respect to a collection of functions)

We say that a scientist \mathcal{M} $PRED_\pi^\beta$ -identifies the set S of total functions if \mathcal{M} $PRED_\pi^\beta$ -identifies each function of S . By $PRED_\pi^\beta$ we denote the class of sets $PRED_\pi^\beta$ -identifiable and by $PRED_\pi$ we denote the class of sets $PRED_\pi^\beta$ -identifiable for some β , i.e., $PRED_\pi = \{S : S \in PRED_\pi^\beta \text{ for some advice } \beta\}$.

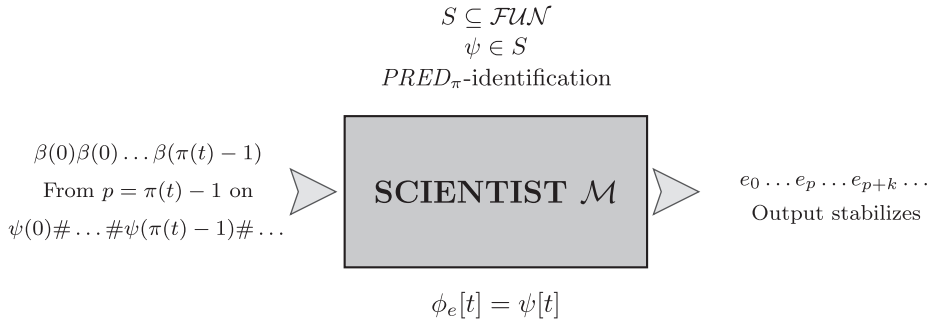


FIGURE 9. For all $t \geq p$, scientist \mathcal{M} on input $\psi[\pi(t)]$ outputs a code e of $\tilde{\psi}$ such that ψ and $\tilde{\psi}$ coincide in the interval $[0, t[$.

DEFINITION 4.5 (Success with respect to a collection of timescales)

For each class of timescales \mathcal{F} , we define $\text{PRED}_{\mathcal{F}}^\beta = \cup_{\pi \in \mathcal{F}} \text{PRED}_\pi^\beta$. We also define $\text{PRED}_{\mathcal{F}} = \{S \in \text{PRED}_{\mathcal{F}}^\beta : \beta \text{ is an advice}\}$.

With Definitions 4.3 and 4.4, we introduce a new concept of function identification: (a) *EX*-identification would correspond to a single hypothesis (over β) produced by the scientist \mathcal{M} , after finitely many observations, (b) *BC*-identification would correspond to a possibly infinite sequence of possibly different correct hypothesis (over β) produced by the scientist \mathcal{M} , after finitely many hypothesis, and (c) *PRED* $_\pi$ -identification corresponds to a possibly infinite sequence of possibly different correct forecasts (over β) for a given timescale.

When a given scientist \mathcal{M} works with a timescale π on input $\psi[\pi(t)]$, using the advice sequence β , his/her output is $\mathcal{M}(\psi[\pi(t)], \beta[\pi(t)])$. In other words, for $t > p$, for $0 \leq x \leq t - 1$, we have

VARYING LAW OF PHYSICS:

$$\phi_{\mathcal{M}(\psi[\pi(t)], \beta[\pi(t)])}[t] = \psi[t] .$$

The scientist \mathcal{M} works as a predictive extrapolation filter and represents for us the concept of a scientific theory ‘buried’ in \mathcal{M} . It captures the idea that a scientific theory is the finite description of some regularity in Nature, although Nature might not be regular enough at that timescale of observations.

Note that a timescale with a small amount of extra information when compared with the variable t , may have very long plateaux, such like $\log^* \in \log^\omega$. This function grows as follows: for values of t such that $17 \leq t \leq 65,536$, the function as value 4. For the next $2^{20,000}$ values of t , starting with 65,537, the function has the same value 5. However, we have that $\log^*(n) \rightarrow \infty$ as $n \rightarrow \infty$. A scientist working with a timescale of this family requires a few bits to compute a ‘slightly’ non-computable function.

That *PRED* $_\pi^\beta$ -identifiable functions exist and that such algorithmic content can be extracted from them is shown below. As we shall see, in relaxing the concept of *scientific theory* discussed by John Case in [9], scientific theories can be independent of whether reality is computable or not, capturing only the observed regularity.

```

FUNCTION  $f_\beta$  :
Function  $f_\beta(x : \mathbb{N}) : \mathbb{N}$ ;
  Function  $\beta(x : \mathbb{N}) : \text{boolean}$ ;
Var  $i, j : \mathbb{N}$ ;
  Begin
     $i := 1$ ;
     $j := 1$ ;
    While  $i \leq x$  Do Begin
       $i := 2^{j+1} - 1$ ;
       $j := j + 1$ 
    End;
     $\beta(j - 1)$ 
  End
    
```

FIGURE 10. Definition of a non-computable function with algorithmic content from a possible pure random sequence β . The timescale is log.

Some subsets of \mathcal{R}_π^β can be exhibited that display similarities with the corresponding sets of learning theory (see [22]): (a) the set of *log-self-describing functions*

$$\mathcal{SD}_{\log}^\beta = \{ \psi \in \mathcal{R}_{\log}^\beta : \phi_{\psi(0)}^\beta = \psi \},$$

(b) the set of *almost everywhere log-self-describing functions*¹¹

$$\mathcal{S}_{\log}^\beta = \{ \psi \in \mathcal{R}_{\log}^\beta : (\text{A}t) \phi_{\psi(t)}^\beta = \psi \}.$$

PROPOSITION 4.6 (Non-emptiness of $PRED_\pi^\beta$)
 $PRED_\pi^\beta$ is not empty.

PROOF. Consider the function β and timescale function log. The new function f_β defined in Figure 10 is in $PRED_\pi^\beta$. Function f_β has the form:

$$\underbrace{\beta(0)}_{1 \times} \underbrace{\beta(1)\beta(1)}_{2 \times} \underbrace{\beta(2)\beta(2)\beta(2)\beta(2)}_{4 \times} \underbrace{\beta(3)\beta(3)\dots\beta(3)}_{8 \times} \underbrace{\beta(4)\beta(4)\dots\beta(4)\dots}_{16 \times}$$

■

PROPOSITION 4.7 (Non-emptiness of \mathcal{S}_{\log}^β)
 \mathcal{S}_{\log}^β is not empty.

PROOF. Consider the function β and timescale log. Let the function $\xi^\beta(e, x)$ be defined as in Figure 11 and let $e \in \mathbb{N}$ be a number such that $\phi_e^\beta(x) = \xi^\beta(e, x)$. The function ϕ_e^β is in \mathcal{S}_{\log}^β . Function ξ^β has the form:

$$eee \underbrace{\beta(1)}_{(2^1-1) \times} \underbrace{e\beta(2)\beta(2)\beta(2)e\beta(3)\dots\beta(3)}_{(2^2-1) \times} \underbrace{e\beta(4)\beta(4)\dots\beta(4)\dots}_{(2^4-1) \times}$$

■

```

FUNCTION  $\xi_\beta$  :
Function  $\xi^\beta(e, x : \mathbb{N}) : \mathbb{N}$ ;
Var  $i, j : \mathbb{N}$ ;
  Begin
     $i := 0$ ;
     $j := 0$ ;
    While  $i < x$  do begin
       $i := 2^j$ ;
       $j := j + 1$ 
    End;
    If  $x < i$  Then Return  $f_\beta(x)$  Else If  $x = i$  Then Return  $e$ 
  End

```

FIGURE 11. Definition of a non-computable self-describing function with algorithmic content from a possible pure random sequence β . The timescale is log.

The following proposition is straightforward.

PROPOSITION 4.8

Both $\mathcal{SD}_{\log}^\beta$ and \mathcal{S}_{\log}^β are $PRED_{\log}^\beta$ -identifiable.

PROOF. The scientist that, on input $\psi[t]$ and $\beta[t]$, with $\psi \in \mathcal{FUN}$ and being β the advice, outputs $\psi(0)$ $PRED_{\log}^\beta$ -identifies $\mathcal{SD}_{\log}^\beta$.

A scientist \mathcal{M} that $PRED_{\log}^\beta$ -identifies the set \mathcal{S}_{\log}^β is defined in the following way: on input $\psi[t]$, with $\psi \in \mathcal{FUN}$, the scientist \mathcal{M} runs in some order $|\psi[t]|$ steps of all the Turing machines of code $\psi(0), \psi(1), \dots, \psi(\log(t)-1)$ on all inputs $0, 1, \dots, \log(t)-1$, being helped by the advice β . The scientist outputs the first code $\psi(i)$ (for some $i < \log(t)$) of the machine that first completes the task of predictive identification $\phi_{\psi(i)}^\beta[t] = \psi[t]$. If such a code is not found after $|\psi[t]|$ steps of each run, then the scientist outputs 0. The scientist \mathcal{M} , for sufficiently large t , always finds the code of ψ relative to β if $\psi \in \mathcal{S}_{\log}^\beta$. ■

5 Characterisation of predictive scientists

The class BC (see Section 2.2) can now be seen as a subclass of $PRED_\pi$, for all π .

PROPOSITION 5.1 ($PRED_\pi$ generalises BC)

If a set S of total functions of type $\mathbb{N} \rightarrow \mathbb{N}$ is BC -identifiable, then S is $PRED_\pi^\beta$ -identifiable, for all timescales π and all advice sequences β .

PROOF. Let $S \in BC$ be witnessed by the computable scientist \mathcal{M} , and let the arbitrary timescale be π . We conclude that, for each function $\psi \in S$, there exists an order p such that, for $t \geq p$, \mathcal{M} on input $\psi[t]$ outputs the programme $\mathcal{M}(\psi[t])$ such that $\phi_{\mathcal{M}(\psi[t])} = \psi$. Since π is surjective and non-decreasing, we conclude that, for each function $\psi \in S$, there exists an order p' such that, for $t \geq p' \geq p$, $\pi(t) > p$, \mathcal{M} on input $\psi[\pi(t)]$ outputs a correct code $\mathcal{M}(\psi[\pi(t)])$ for ψ . As a consequence, we also have $\psi[t] = \phi_{\mathcal{M}(\psi[\pi(t)])}[t]$.

¹¹Self-describing infinitely many times. The quantifier \mathcal{A} on a variable t has the meaning ‘for infinitely many values of t ’.

Let \mathcal{M}' be the $PRED_{\pi}^{\beta}$ -scientist such that, on input $\psi[x]$, ignores the advice function β and calls the scientist \mathcal{M} on $\psi[x]$. Consequently, for all $t \geq p'$, we have $\phi_{\mathcal{M}'(\psi[\pi(t)], \beta[\pi(t)])}[t] = \phi_{\mathcal{M}(\psi[\pi(t)])}[t] = \psi[t]$. We conclude that $S \in PRED_{\pi}^{\beta}$, since to predict the values of $\psi[t]$, we only need input of size at most $\pi(t) \geq p$. ■

We consider the families $\log^{(\alpha)}$, for each ordinal α , as the canonical families of reasonable timescales and define:

DEFINITION 5.2 ($PRED_{\alpha}$)

Let β be some advice sequence. We say that a set S of (total) functions of type $\mathbb{N} \rightarrow \mathbb{N}$ is in $PRED_{\alpha}^{\beta}$, for some ordinal α , if S is $PRED_{\log^{(\alpha)}}^{\beta}$ -identifiable.¹²

PROPOSITION 5.3 (\mathcal{R} is in the inductive limit)

There exists an advice β such that $\mathcal{R} \in PRED_{\alpha}^{\beta}$, for all ordinal $\alpha \in \omega^{\omega}$.

PROOF. Let $\psi \in \mathcal{R}$. Just consider the scientist \mathcal{M} that, relatively to some advice function β to be specified, on input $\psi(0)\#\dots\#\psi(\pi(t)-1)\#$, searches for the code of a computable function in \mathcal{R} and outputs the least code e such that $\phi_e[\pi(t)] = \psi[\pi(t)]$. Let the ω -sequence β be defined as follows: for all $i \in \mathbb{N}$, $\beta(i) = 1$ if and only if the i -th partial computable function is defined for all inputs. For each $\psi \in \mathcal{R}$, only a finite number p of bits of β are enough to identify ψ . Once the scientist finds p , the values of $\beta(i)$, for $i > p$, are not needed, since the scientist converges on e .

We conclude that $\mathcal{R} \in PRED_{\alpha}^{\beta}$, for all ordinals $\alpha \in \omega^{\omega}$, since to identify each function $\psi \in \mathcal{R}$ only a fixed number of bits of the above β are needed. ■

A computable phenomenon, described by some function in \mathcal{R} , needs only a finite number of advice bits to be identified. A non-computable phenomenon needs an unbounded number of advice bits to be identified. The degree of regularity of a phenomenon translates into the degree of growth of the timescale. The timescale grows less as ‘less computable’ the law is. We can define a degree of regularity of a law of Nature by means of the family of timescales (see Section 3).

DEFINITION 5.4 (Degree of predictability of a physical law)

A physical law presented by the set S of all its instances is said to be of ordinal degree α if $S \in PRED_{\alpha}$ but $S \notin PRED_{\alpha+1}$.

This definition suggests that very high degrees of regularity do not prevent the non-computable character of the physical law. It just tells us that we can live with the ‘illusion’ of certainty of the regularity in Nature.

PROPOSITION 5.5 (From non-computable to computable)

If a set S of total functions of type $\mathbb{N} \rightarrow \mathbb{N}$ is such that $S \in PRED_{\pi}^{\beta}$, for some computable advice sequence β , then each $\psi \in S$ is a computable function and $S \in BC$.

PROOF. Let β be a computable sequence and \mathcal{M} a computable scientist that $PRED_{\pi}^{\beta}$ -identifies $\psi \in S$. Let T be the canonical text for ψ . The routine in Figure 12 computes the function φ . Applying the s_n^m -theorem to the function φ , we get $\phi_{s(\sigma)}(x) = \varphi(\sigma, x)$, for some computable function s . Thus $\phi_{s(T[1])}, \dots, \phi_{s(T[t])}, \dots$ is an infinite enumeration of a finite number of (partial) computable functions, since the scientist \mathcal{M} , on input the text $T[\pi(t)]$, with computable advice β , for $t \geq p$, for some p ,

¹²However, we take $PRED_0 = \mathcal{F}\mathcal{U}\mathcal{N}$, notation that is consistent with Definition 4.3.

```

FUNCTION  $\varphi$  :
Function  $\varphi(\sigma : DATA; x : \mathbb{N}) : \mathbb{N}$ ;
  Function  $\beta(t : \mathbb{N}) : Boolean$ ; %  $\beta$  is computable
Var  $i, j, e : \mathbb{N}$ ;
Begin
  While  $i < x$  Do Begin
     $j := |\sigma| - 1$ ;
     $e := \mathcal{M}(\sigma, \beta[|\sigma|])$ ;
    For  $k := j + 1$  To  $\pi^{-1}(j)$  Do  $\sigma := \sigma \diamond \langle k, \phi_e^\beta(k) \rangle$ ;
     $i := \pi^{-1}(j)$ 
  End;
Return  $\hat{\sigma}(x)$ 
End

```

FIGURE 12. A function φ that recursively extrapolates ψ , identified by the prefix σ , over the whole set of natural numbers.

```

 $\varphi$  COMPUTES  $\beta$  :
Function  $\varphi(\nu : \{0, 1\}^*; t : \mathbb{N}) : \mathbb{N}$ 
Var  $e : \mathbb{N}; \mu : \{0, 1\}^*$ ;
Begin
   $\mu := \nu$ ;
  While  $|\mu| < t$  Do Begin
     $e := \mathcal{M}(\mu, \mu)$ ;
     $\mu := \{e\}(0) \# \dots \# \{e\}(\pi^{-1}(|\mu| - 1)) \#$ 
  End While;
Return  $\hat{\mu}(t)$ 
End

```

FIGURE 13. Computing the bits of β .

outputs codes of the same function. The (total) computable function ψ occurs in such a list (it is the last repeated element) and is, therefore, computable and *BC*-identified. ■

The physical law appear in our framework as a set of non-computable functions S , the instances of the physical law, where each of the functions is the kind of a ‘convolution’ of a more or less random sequence β with a computable function, in the sense that the functions in S can be described as functions of \mathcal{R}_π^β .

PROPOSITION 5.6 (Recursive β)

If β is $PRED_\pi^\beta$ -identifiable for some strictly sublinear π , having β as advice, then β is total computable.

PROOF. Let π be the timescale and \mathcal{M} the scientist witnessing $\beta \in PRED_\pi^\beta$. On input $\beta[\pi(t)]$, from some order $t \geq p$, we have

$$\phi_{\mathcal{M}(\beta[\pi(t)], \beta[\pi(t)])}^\beta[t] = \beta[t].$$

If we know the prefix $\beta[t]$ of β , we can predict $\pi^{-1}(t-1)$ bits of β . The whole job can be performed by function φ given in Figure 13. (For sufficient large number of observations, the sequence ν

<div style="border: 1px solid black; padding: 2px; margin-bottom: 5px; text-align: center;"> COMPUTING $g(n)$ BITS OF β </div> <pre> Procedure Begin Input n; $e := \mathcal{M}(\psi[f(n)], \gamma[f(n)]);$ % ϕ_e provides $g(n)$ bits, $\beta_0, \dots, \beta_{g(n)-1}$, of β $\mu := \lambda$; % Add 1, read in binary and remove the bit 1 $z := 0$; For $i := 1$ To n Do Begin If $g(i) \neq z$ Then Begin % Produces the next bit of β $\mu := \mu \diamond$ "the $g(i)$-th bit of $(\phi_e[n] + 1)_2$"; $z := g(i)$; End If; End For; Output μ % The μ is indeed a prefix of the true advice β End. </pre>
--

FIGURE 14. Establishing a hierarchy.

is a ‘locking sequence’ for \mathcal{M} .) Then there exists a total unary computable function s such that $\phi_{s(v)}(t) = \varphi(v, t)$. For some order $t \geq |v| \geq p$ on, we have $\beta(t) = \phi_{s(v)}(t) = \varphi(v, t)$. ■

6 The existence of a non-collapsing hierarchy of predictive scientists

We now prove that scientists can be hierarchized according to their intrinsic timescales, i.e., long-term predictive scientists ‘compute’ less functions than short-term predictive scientists. The hierarchy of predictive scientists reflects the hierarchy of timescales.

PROPOSITION 6.1

If \mathcal{F} and \mathcal{G} are classes of timescales such that $\mathcal{F} < \mathcal{G}$, then there is a set in $PRED_{\mathcal{G}}$ that is not in $PRED_{\mathcal{F}}$.

PROOF. We prove that if there is a $g \in \mathcal{G}$ so that for all $f \in \mathcal{F}$ we have $f \in o(g)$, then there is a set in $PRED_{\mathcal{G}}^{\beta}$, for some advice β , that does not belong to $PRED_{\mathcal{F}}^{\gamma}$, for any advice γ . We assume that β is in $\mathcal{K}[\lambda t.t]$ but not in $\mathcal{K}[\lambda t.\eta t]$ for some $\eta < 1$. It will be enough to show that there is a $PRED_{\mathcal{G}}^{\beta}$ -identifiable function ψ which is not $PRED_{\mathcal{F}}^{\gamma}$ -identifiable, for any advice γ . We define the function ψ in the following way:

$$\psi(t) = (1(\beta \downarrow_{g(t)})0^{t-g(t)})_{10} - 1$$

This function ψ belongs trivially to the class of $PRED_{\mathcal{G}}^{\beta}$. We prove that the same function ψ does not belong to $PRED_{\mathcal{F}}^{\gamma}$, for any advice γ . Suppose that some scientist \mathcal{M} with advice $\gamma \downarrow_{f(n)}$ identifies the function ψ up to n values. A scientist \mathcal{M}' can call the scientist \mathcal{M} on $\langle \psi[f(n)], \gamma[f(n)] \rangle$ to compute $g(n)$ bits of β the procedure of Figure 14.

Since $f \in o(g)$, then, for all but finitely many t , $f(n) < \eta g(n)$, for arbitrarily small η , meaning that we can compute, for all but finitely many t , $g(n)$ values of β using a guess of length $\eta g(n)$, for arbitrarily small η , contradicting the fact that $\beta \notin \mathcal{K}[\lambda t.\eta t]$. ■

M OUTPUTS CODE OF PROGRAMME :

```

Function  $\varphi(x : \mathbb{N}) : \mathbb{N}$ ;
Const  $\mu : \text{list of } \mathbb{N} \times \mathbb{N}$ ;
Begin
  Case  $x$  of
     $x \in \text{dom}(\hat{\mu}) : \text{Return } \hat{\mu}(x)$ 
  End;
  Return 0
End
    
```

FIGURE 15. Sending the non-converging code of the full list of inputs so far seen.

PROPOSITION 6.2 (*PRED*-hierarchy)

For all ordinal $\alpha \in \omega^\omega$,

$$EX \subset BC \subset \dots \subset \text{PRED}_{\alpha+1} \subset \text{PRED}_\alpha \subset \dots \subset \text{PRED}_0$$

PROPOSITION 6.3

Every set $S \subseteq \mathcal{FUN}$ is $(\text{PRED}_0^\beta \Rightarrow \text{PRED}_{\lambda,t}^\beta)$ -identifiable for every guess β .

PROOF. Every function of type $\mathbb{N} \rightarrow \mathbb{N}$ can be reproduced from itself without any guess sequence β . Take as $\beta : \mathbb{N} \rightarrow \{0, 1\}$ any arbitrary ω -sequence and as order $p = 0$. We have that, for every $t \geq p$, the scientist \mathcal{M} such that, on input $\psi[\pi(t)] = \psi[t]$, outputs the code of the function φ given by Figure 15, is such that $\phi_{\mathcal{M}(\psi[t], \beta[t])}[t] = \psi[t]$. ■

DEFINITION 6.4

A function ψ is unpredictable if it is $\text{PRED}_{\lambda,t}$ -identifiable but not PRED_π -identifiable for any timescale π in any smaller class.

PROPOSITION 6.5 (Based on Kobayashi [23], Loveland [24])

The characteristic functions of recursively enumerable sets belong to PRED_{\log^2} .

PROOF. We prove that each recursively enumerable set belongs to the class $\text{PRED}_{\log^2}^\beta$, for a suitable β . We first follow Loveland in [24]. Let A be a recursively enumerable set and χ_A its characteristic sequence. For each $n \in \mathbb{N}$, let the (possibly) non-computable function $h_A : \mathbb{N} \rightarrow \mathbb{N}$ be defined as $h_A(n) = \#\{m \in \mathbb{N} : m \leq n \wedge \chi_A(m) = 1\}$. In the general case, to compute h_A , for each input n , an advice of size $\lceil \log(n) \rceil + c$ (e.g., to provide $h_A(n)$ in binary) is enough, where c is the size of a Turing machine \mathcal{M} that enumerates A . (If the set A is finite, then the (computable) advice becomes of constant size, coinciding with the cardinal of A .) A programme to predict the bits of χ_A can be devised in the following way: on input n , it makes up a list of n 0s, runs the enumerator \mathcal{M} , and keeps replacing the 0s of the list by 1s at the appropriate positions, whenever one such 1 is found, until exactly m 1s are counted. In the end, the machine outputs the sequence of the first n bits of χ_A . For this purpose, an advice of size $\lceil \log(n) \rceil + c$ is given in the input to the universal Turing machine, paired with n .

However, considering Definition 3.10 of Kolmogorov complexity of infinite sequences and Definition 4.3 of predictive scientist, some more work has to be done, since a unique (infinite) sequence β has to be provided to the scientist to identify the given recursively enumerable sequence. We notice that a sequence of n bits can be successively approached by $\lceil \log(n) \rceil$ sequences of sizes $\lceil \log(2^0) \rceil, \lceil \log(2^1) \rceil, \lceil \log(2^2) \rceil, \dots, \lceil \log(2^m) \rceil$, where m is the least number such that $2^m \geq n$. For each such prefix of size 2^k , the corresponding prefix $\chi_A[2^k]$ needs advice of size $k \in \mathcal{O}(\log(n))$. All the

advices can now be concatenated under suitable encoding to produce a sequence $\beta[\lceil \log(n) \rceil^2]$ that will be enough to predict the bits of $\chi_A[n]$. Moreover, to predict a sequence χ_A , the scientist just has to run the enumerator of A on help by the advice β . ■

From this proof, we see that Kolmogorov complexity goes well with learning theory, for while the latter predicts future values of functions, the former bounds the queries to the advice sequence. Note that, from Proposition 5.3, we can predict *all* the characteristics χ_A of recursively enumerable sets with just the same β . However, such a proof does not bound the number of bits of β needed for an input sequence $\chi_A[n]$ of size n , although it tells us that for sufficiently large n it becomes constant. Moreover, although Proposition 6.5 is not that much useful, in the sense that we have to have one scientist for each recursively enumerable set, it shades some light on the difference between the learning paradigm and the Kolmogorov complexity paradigm.

7 Conclusions

We considered that the laws Physics express only computable relations between the diverse physical concepts. Either as more observations are made or as the precision of measurements increase, Physics might undergo some change. For example, as the precision of measurements increases, the orbit of a planet (e.g. Mercury) is first explained as a consequence of Newtonian theory of gravitation and, two centuries after, as a consequence of Einstein (linearized) field equations. For example, as the universe expands, the ‘universal constants’ may vary altering the Physics. According to yet another perspective, such as that of Peirce (see [25]), the world does not obey exactly to fixed laws but exhibits a regularity that is captured by possibly varying laws.

7.1 Laws of Physics might not be computable

We presented a model of the still learnable evolving physical law and a classification of its regularity.

Learning theory provides also an explanatory framework of concepts occurring in the Philosophy of Science. However, it does not fit exactly with the character of the empirical laws of science. One of the reasons is that most of the empirical laws of science we can imagine are described by primitive recursive functions, as explained above. However, the limits of learnability and the richness of the theory only manifest themselves when the realm of scientific laws turn to be computable but not primitive recursive. Non-primitive recursive laws did not have an expression in the elementary laws of natural science.

Modern Physics also expresses the laws of Nature by computer programmes (possibly inputting random bits of some source) that possibly produce non-primitive recursive behaviours. The possibility of non-computable behaviours also encourage the development of the proposed approach.

The model becomes more expressive if we think on the laws of Nature that, instead of being described by elementary functions of Analysis or by differential equations integrable by elementary functions, are described by computer programmes such like those of the so-called complex systems, where emergent properties are not derived from the properties of their components but rather emerge as a result of their interacting parts. Such systems that are not being analysed by reductionist theorists but instead by holistic top-down methods may, indeed, exhibit behaviour that can not be identified by *EX* or even *BC* scientists. Instead of a differential equation, the system is explained and forecast from scratch by a computer programme implementing explanatory top-down rules. These scientific laws are closer to the analysis done in this article.

The reader may consider that we largely avoid problems of abstraction and focus on Physics as an activity based on observed everyday practice (e.g., see Floridi’s book [19]). But, if we think further

in the development of a theory of Physics such as Maxwell's Electromagnetic Theory, we realize that the (theoretical) physicist (and not the empiricist), instead of looking at a quiz of numbers or relations (such as 'Boyle') to find the next empiric law of Physics, looks at samples of axioms/theorems such as Coulomb's law of electrostatics, Ampère's law for electric circuits, Faraday's law of induction, etc., and by adding some new hypothesis, such as displacement current, establishes his/her theory. The scientist instead of a quiz of numbers or relations is confronted with a quiz of theorems and looks for a theory. However, this apparent change of level of abstraction finds its automorphic character in the fact that (for simplicity!) searching for the least code of a theory (over some grammar of formulas) that produces the theorems so far seen, so far proved, is, in a sense, equivalent to searching for the least code number of a set.

7.2 *Future work*

We shortly address some interesting ideas to be analysed.

Learning procedures can be compared to Maupertuis or Fermat principles, i.e., to teleological principles in Physics. For example, let us examine a model of learnability of primitive recursive functions. Starting with a prefix of the graph of a primitive recursive function ψ , we search for the least code $e \in \mathbb{N}$ that satisfies the equation $\phi_e[t] = \psi[t]$. We repeat the procedure for increasing values of t until it converges to the right code. In the middle we have some problem to solve... since some of the codes are of functions that do not halt for some inputs. For the primitive recursive function case, however, there is an encoding into the natural numbers that will do, and such an *initial condition* e may be found. In the process of finding the right initial condition e , finitely many prefixes of the graph of each function have to be analysed before the process converges into e . In Physics, once we know a prefix of the trajectory of the system (a finite prefix of its topology), we may want to find the initial condition to compute and forecast the entire trajectory. We apply the so-called inverse methods: from the observations and the known laws of Physics we compute an approximation to the initial condition that is minimal according with some principle; then we compute the presumable trajectory and compare it with the observed trajectory; then we repeat the procedure getting better approximations to the initial condition converging to the desired accuracy.

A second interesting aspect of our framework is the study of strong and weak causality in Physics taking into consideration the degrees of predictability of the *physical law*. In chapter 10 of his book on causality (see [7]), Mario Bunge addresses the lawfulness of historical processes in a way that can be modelled as a predictive scientist with some free spirit:

Positivists have demanded that the concept of causation be replaced by that of law; they have required that science forsake what Comte called '*la veine recherche des causes*'; and they have imagined that science has already eliminated the common sense notion of causal influence, 'substituting the entirely different *denkmittel* of 'law'. Some antipositivist, on the other hand, have held that lawfulness is insufficient, a causal connection having to be sought *behind* every law. An examination of the nature of scientific law has shown us, however, that science has followed neither advice, since it contains both causal and noncausal laws, as well as laws in which various categories of determination concur.

Acknowledgements

The talks of Barry Cooper were the pivot to develop this framework, for we wanted to understand possible relations between degrees of computability and Physics. We thank John Case for having

introduced us to Learning Theory. The proposed model was influenced by my work with Edwin Beggs and John V. Tucker in [5] on a ‘kind of’ timescales and by the papers [1, 2] on applications of Kolmogorov complexity to classify real numbers according to their compressibility. The research of José Félix Costa is supported by Fundação para a Ciência e Tecnologia, PEst – OE/MAT/UI0209/2011.

References

- [1] J. L. Balcázar, R. Gavaldà, and M. Hermo. Compressibility of infinite binary sequences. In *Complexity, Logic, and Recursion Theory*, Vol. 187 of *Lecture Notes in Pure and Applied Mathematics*, A. Sorbi, ed., pp. 1175–1183. Marcel Dekker, Inc., 1997.
- [2] J. L. Balcázar, R. Gavaldà, H. Siegelmann, and E. D. Sontag. Some structural complexity aspects of neural computation. In *Proceedings of the Eighth IEEE Structure in Complexity Theory Conference*, pp. 253–265. IEEE Computer Society, 1993.
- [3] J. L. Balcázar and M. Hermo. The structure of logarithmic advice complexity classes. *Theoretical Computer Science*, **207**, 217–244, 1998.
- [4] J. Bārzdiņš, ed. Two theorems on the limiting synthesis of functions. In *Theory of Algorithms and Programs*, vol. I, pp. 82–88. Latvian State University, 1974.
- [5] E. Beggs, J. F. Costa, B. Loff, and J. V. Tucker. Oracles and advice as measurements. In *Unconventional Computation (UC 2008)*, Vol. 5204 of *Lecture Notes in Computer Science*, C. S. Calude, J. F. Costa, R. Freund, M. Oswald, and G. Rozenberg, eds, pp. 33–50. Springer, 2008.
- [6] L. Blum and M. Blum. Toward a mathematical theory of inductive inference. *Information and Control*, **28**, 125–155, 1975.
- [7] M. Bunge. *Causality and Modern Science*. Dover, Third Revised Science, 1979.
- [8] J. Case. Directions for computability theory beyond pure mathematical. In *Mathematical Problems from Applied Logics. New Logics for the XXIst Century. II.*, D. Gabbay, S. Goncharov, and M. Zakharyashev, eds, pp. 53–98. Springer, 2006.
- [9] J. Case. Algorithmic scientific inference, within our computable expected reality. *International Journal of Unconventional Computing*, **8**, 192–206, 2012.
- [10] J. Case, S. Jain, and S. N. Manguelle. Refinements of inductive inference by Popperian and reliable machines. *Kybernetika*, **30**, 23–52, 1994.
- [11] J. Case and S. N. Manguelle. Refinements of inductive inference by Popperian machines, 1979. Technical Report 152, SUNY/Buffalo.
- [12] J. Case and C. Smith. Anomaly hierarchies of mechanized inductive inference. In *Proceedings of the 10th Annual ACM Symposium on Theory of Computing, May 1-3, 1978*, R. J. Lipton, W. A. Burkhard, W. J. Savitch, E. P. Friedman, and A. V. Aho, eds, pp. 314–319. ACM, 1978.
- [13] J. Case and C. Smith. Comparison of identification criteria for machine inductive inference. *Theoretical Computer Science*, **25**, 193–220, 1983.
- [14] B. Cooper and P. Odifreddi. Incomputability in Nature. In *Computability and Models, Perspectives East and West*, University series in mathematics, B. Cooper and S. Goncharov, eds, pp. 137–160. Springer, 2003.
- [15] B. S. Cooper. Mathematics, metaphysics and the multiverse. In *Computation, Physics and Beyond 2012, Lecture Notes in Computer Science*, M. J. Dinneen, B. Khossainov, and A. Nies, eds, pp. 252–267. Springer, 2012.
- [16] B. S. Cooper. Turing’s titanic machine? *Communications of the ACM*, **55**, 74–83, 2012.
- [17] R. Daley. On the error correcting power of pluralism in BC-type inductive inference. *Theoretical Computer Science*, **24**, 95–104, 1983.

- [18] J. Feldman. Some decidability results on grammatical inference and complexity. *Information and Control*, 20:244–262, 1972.
- [19] L. Floridi. *The Philosophy of Information*. Oxford University Press, 2013.
- [20] E. M. Gold. Language identification in the limit. *Information and Control*, **10**, 447–474, 1967.
- [21] C. G. Hempel. Fundamentals of concept formation in empirical science. In *Foundations of the Unit of Science: Toward an International Encyclopedia of Unified Science*, O. Neurath, R. Carnap and C. F. W. Morris, eds, Vol. 2, pp. 652–745. University of Chicago Press, 1970.
- [22] S. Jain, D. N. Osherson, J. S. Royer, and A. Sharma. *Systems That Learn. An Introduction to Learning Theory*, 2nd edn., The MIT Press, 1999.
- [23] K. Kobayashi. On compressibility of infinite sequences. *Technical Report C-34, Research Reports on Information Sciences*, 1981.
- [24] D. W. Loveland. A variant of the Kolmogorov concept of complexity. *Information and Control*, **15**, 115–133, 1969.
- [25] C. S. Peirce. *The Logic of Interdisciplinarity*. The Monist Series, E. Bisanz, ed., Akademie Verlag, 2009.
- [26] K. R. Popper. *The Logic of Scientific Discovery*. Routledge, 1935.

Received 9 March 2013