

A Stochastic Complexity Perspective of Induction in Economics and Inference in Dynamics

K. Vela Velupillai*
Department of Economics
National University of Ireland
Galway, Ireland
&
Girton College
Cambridge CB3 0JG
England

Prepared for the:
FESTSCHRIFT IN HONOUR OF JORMA RISSANEN
(On the Occasion of his 75th Birthday)

November 21, 2007

*I am very particularly indebted to Duncan Foley, Francesco Luna, John McCall, Shu-Heng Chen and Stefano Zambelli. They encouraged my early and continuing interest in Jorma Rissanen's work from their own particular viewpoints of Bayesian statistics, Gold's model of learning, de Finetti's theory of probability, *MDL* and Chaitin's algorithmic information theory, respectively. However, they are not responsible for any of the remaining infelicities. My own view was, from the outset, shaped by my determination to fashion the subject of computable economics on the foundations of recursion theoretic and constructive mathematics - i.e., on the works of Turing, Kolmogorov, Brouwer and Bishop. Hence, I came to stochastic complexity, as Rissanen originally did, from 'Kolmogorov complexity' theory.

Abstract

Rissanen's fertile and pioneering *minimum description length* principle (*MDL*) has been viewed from the point of view of statistical estimation theory, information theory, as stochastic complexity theory¹ – i.e., a computable approximation to Kolmogorov Complexity – or Solomonoff's recursion theoretic induction principle or as analogous to Kolmogorov's sufficient statistics. All these – and many more – interpretations are valid, interesting and fertile. In this paper I view it from two points of view: those of an algorithmic economist and a dynamical system theorist. From these points of view I suggest, first, a recasting of Jevons's sceptical vision of induction in the light of MDL; and a complexity interpretation of an undecidable question in dynamics.

¹I am using 'stochastic complexity' in a kind of 'generic' way. Rissanen has, over the past three decades, gradually refined the exact formal meaning of the phrase and I believe his most mature views are now represented in [1]. The kind of meaning I have in mind is what I learned from Rissanen's early writings on MDL, for example in [24], p.1080, emphasis in the original:

"[I]f ... a shortest description of the data, to be called STOCHASTIC COMPLEXITY is found in terms of the models of a selected class, there is nothing much further anyone can teach us about the data; we know all there is to know."

1 A Personal Preamble

To paraphrase the famous reply of Laplace to Napoleon, who wondered why the word 'God' did not appear in *Mécanique Céleste*, we could state that 'the assumption of a 'true' distribution is not needed in this theory'.

Jorma Rissanen

I have shared many moments of intellectual and personal splendour with Jorma Rissanen. One serendipitous conjunction relates to my first published, technical, article, which was in 1978, in Volume 14 of *Automatica*, [27]. I did not, of course, know, then, that Jorma Rissanen's first published, pioneering, paper on *stochastic complexity* – or, the *Minimum Description Length* principle (henceforth, *MDL*) – was also in that same volume of the same *Journal*²! Justifiably, that paper on *MDL* spawned a path-breaking research program that has, in one strand, developed into *Algorithmic Statistics*. It gives me great and undiluted pleasure to state that my own, much humbler piece, in that same volume of *Automatica*, was also the fountainhead for what I have developed into the research program on *Algorithmic Economics*!

In the intervening 30 years, particularly in its second half, I have had the pleasure and privilege of hosting Jorma Rissanen at numerous venues, exotic and otherwise, trying to make his fertile and fascinating research program more familiar to obdurate economists. I believe a measure of success can now be seen, albeit taking place at snail's pace.

I had read, quite by chance (sic!), an expository piece on *stochastic complexity* in an issue of the *IBM Research Magazine*³ around the time I was trying to establish the *Center for Computable Economics (CCE)*, in the department of economics, at UCLA, in the academic year 1990-91. The modest initial success, together with funds for a seminar series on Computable Economics, gave me the chance to invite Jorma Rissanen to give a talk at the *CCE* seminar series, as one of its first speakers, in autumn, 1991. Soon after that I organised a 'Summer School' in Computable Economics, in July, 1992, sponsored by Aalborg University in Denmark, at the beautiful *Dronninglund Slot* in Nordjylland. Naturally, Jorma Rissanen was one of the key speakers at that event.

²Rissanen's classic was published in the September issue of Volume 14, 1978 and mine in the November issue of the same Volume ([23], [27]). Mine had been presented at an IFAC meeting in Vienna the year before. My path towards what I now call Algorithmic Economics began with *computational complexity theory*. I like to think there is a further *serendipity* even here: one strand of the tradition from which Jorma Rissanen created MDL arose from *algorithmic complexity theory*, as is well documented in several articles tracing his thought on these matters.

³The 'expository' piece, by Rowan Dordick of IBM's 'communications department', [7], had an eye-catching title – *Understanding the 'go' of it*, quoting Maxwell – and an attractive blurb, (with a photograph of Jorma Rissanen at the blackboard (those were the days...!)), which said:

"A novel approach to statistical inference – the theory of stochastic complexity – holds that the best description of data is the shortest one. "

Most recently I set up the *Computable and Behavioural Economics Research Axis*⁴ (COBERA) in the department of economics at the National University of Ireland, in Galway. One of the first events sponsored by COBERA was a ‘Spring School’ on Computable Economics, in March, 2005. Naturally, Jorma Rissanen was again one of the key lecturers at this event, too.

In all of the above events the audience was predominantly made up of advanced graduate students, senior and junior faculty and interested outside participants, almost all of whom were economists. However, the distinguished speakers – like Jorma Rissanen – were not all economists; apart from Jorma Rissanen, there were recursion theorists (Piergiorgio Odifreddi, F.A. Doria), algorithmic information theorists (Greg Chaitin), game theorists (Ken Binmore), dynamical system theorists (Ralph Abraham, Joe McCauley), and others, all of whom were united by being motivated by an algorithmic approach to theory and application in the sciences, both pure and applied.

Jorma Rissanen was always a persuasive lecturer and an engaging participant at all of these events. Economists of widely varying persuasions – in statistical methodology and mathematical epistemology – were always fascinated by his wonderful lectures, always prepared with utmost care and delivered with immaculate clarity. On many occasions his talks were interrupted by genuinely perplexed members of the audience who were struggling to absorb a whole new set of concepts with which to understand a fascinating framework and methodology. On occasions there was also one or another famous, but obdurate, economists, entrenched in orthodoxy, who was unable to dissociate himself from the traditional frameworks that shackled his thoughts and practice.

I would like to end this brief personal preamble with a pleasant recollection of an event that I have had occasion to repeat almost every time I have chaired a session where Jorma Rissanen has been the lecturer. On this particular occasion, after Jorma Rissanen’s beautifully crafted lecture on stochastic complexity and statistical estimation, the following brief dialogue occurred between a very distinguished game theorist (referred to as DGE), not known for any competency in statistical methodology, and Jorma (JR):

DGE: (In an irritated tone), ‘You seem to be suggesting that your method is the only one around. You must know that there are many other methods, and some have survived the test of time, too.’

Pin-drop silence in the lecture hall (at Dronninglund Slot).

JR: (In a perfectly calm and conciliatory tone), ‘Oh, I am so sorry; I did not mean to suggest that MDL was the only statistical method around. I do apologise if I gave that impression.’

Pause and continued silence in the lecture hall; not even a whisper or a murmur among the distinguished collections of lecturers and auditors, among whom were very famous economists like Bob Clower, Axel Leijonhufvud, Michael Intrilligator, John McCall; computer scientists like Berc Rustem, Greg Chaitin and Piergiorgio Odifreddi; and so on.

Then, after only a brief pause, which seemed like eternity:

⁴Now defunct.

JR: *'But it is the best [method available]!'*

The whole hall erupted in appreciative and almost unanimous (i.e., except one member of the audience!) laughter and applause.

Jorma Rissanen continues a distinguished Finnish tradition of making Induction a scientifically respectable enterprise, free of the nihilistic scepticism propagated by ill-informed scholars of Hume and Mill⁵, particularly in economics and the philosophy of science. His great predecessors and contemporaries in the rich *Finnish tradition* of the mathematical epistemology of induction are, *among others*, of course Georg Henrik von Wright, Jaako Hintikka, Ilkka Niiniluoto and Risto Hilpinen⁶. In my own economics education at Cambridge in the early 1970s, under the inspiring supervision of Richard Goodwin, I was advised, wisely as it turned out later, to attend the lectures given by Ian Hacking in the philosophy department. Fortunately, Hacking was just then lecturing, broadly, on issues of induction and probability and, of course, the works of Wittgenstein's immediate successor as the Knightbridge Professor of Philosophy at Cambridge, Georg Henrik von Wright, were often brought into focus. Margaret Anscombe was often at those lectures and, occasionally, a brief dialogue took place between Hacking and Anscombe, to which we – students – were privileged auditors.

It is a pleasure and a privilege to pay homage to a pioneer scientist of uncompromising integrity and undiluted personal warmth.

The paper is divided into four subsequent sections. A brief methodological discussion, of lessons learned from Rissanen's modelling philosophy, is the content of the next section. In section 3, the main, substantively economic section of the paper, I try to reinterpret a celebrated sceptical – even hostile – vision of inductive inference by one of the pioneers of modern economic theory from the point of view of MDL. In section 4, motivated by an issue in economic dynamics, I try to pose an undecidable problem in dynamical systems theory as an inference problem and formulate its Kolmogorov complexity. The concluding section 5 consists of speculative thoughts on Algorithmic Economics as a companion in arms of Algorithmic Statistics, Algorithmic Randomness and Algorithmic Information Theory.

⁵I have in mind, in particular, Jevons in economics and Popper in the philosophy of science. What I have to say about Jevons is given in section 3, below; I have had my say on Popper, *from the point of view of MDL* in [30]. Li and Vitanyi quite pungently, but accurately (I think) note, [31], p.448:

"Unsatisfactory solutions [to the problem of scientific inference] have been provided by philosophers like R.Carnap and K.Popper."

I suppose they should have been a little more precise and designated these two worthy individuals as 'philosophers of science'! In any case, my case against Popper from the point of view of MDL, substantiating the Li-Vitanyi claim, is fully described and discussed in detail in [30].

⁶von Wright's magisterial exposition of induction in the probabilistic tradition is in [32] & [33]; for Hintikka's views (and Niiniluoto's), in a Carnapian tradition, the best source may well be, [11] & [12]. A different source for Niiniluoto's work on induction is, of course, his early joint monograph with another distinguished Finn, Tuomela, in, [21]. One reference for Hilpinen's work on inductive logic is [9].

2 Extracting Methodological Precepts for *Algorithmic Economics* from Rissanen's Modelling Philosophy

"Regarding the ultimate model, no algorithmic procedure to find it can exist, as shown in the theory of the *algorithmic complexity*, Solomonoff (1964), Kolmogorov (1965), Chaitin (1973), which also *is the spiritual father of our main notion.*"

Jorma Rissanen, [25], p.224; italics added.

Rissanen's philosophy of stochastic complexity suggests a way of exorcising the search for that traditional 'Will o' the Wisp' in formal modelling exercises: the '*true*' model underpinning observable, empirical data. Secondly, in one of its recent incarnations, the modelling philosophy of stochastic complexity has evolved into *algorithmic statistics*. As defined by the three pioneers, *algorithmic statistics* is the theory of the '*relation* between an individual data sample and an individual model summarizing the information in the data', [8], p. 2443. In this theory the search is for an 'absolute notion' of such a 'relation' in analogy with the way 'Kolmogorov complexity is the accepted absolute measure of information content of an individual finite object' (*ibid*). Thirdly, the concept of universality – either of the Universal Turing Machine in recursion theory, or of the prior in the Solomonoff scheme or of models in Rissanen's recent work on stochastic complexity.

Finally, I want to return to one of the earliest insights and interpretations of 'stochastic complexity' as a computable approximation of the uncomputable Kolmogorov complexity (or, equivalently, of Solomonoff's uncomputable 'universal prior'). The orientation of my own research in algorithmic economics has been almost entirely determined by this particular insight. Therefore, let me, touch on this point, very briefly, before going on to the main sections of the paper.

In his original paper introducing the stochastic complexity approach to statistical inference as inductive inference from finite data sequences, Rissanen acknowledged his indebtedness to Kolmogorov ([18], p. 465). It is generally understood, by scholars who have closely studied the origins and evolution of Rissanen's ideas on stochastic complexity, that this horn of the original motivation – the other being Akaike's *AIC* model⁷ – led to the idea of stochastic complexity being a computable approximation to the uncomputable Kolmogorov complexity. In this original paper, Kolmogorov defined the notion that has forever since

⁷Even with some reservations, Rissanen is handsome in his acknowledgement to Akaike, [25], p.224:

"[W]e are indebted to Akaike's pioneering and innovative work for inspiration in our own efforts."

then been associated with his name in the following way:

$$K_\phi(y|x) = \begin{cases} \min_{\phi(p,x)=y} l(p) \\ \infty \quad \# p \text{ s.t } \phi(p,x) = y \end{cases} \quad (1)$$

where:

$\phi(p,x) = y$: a partial recursive function – the ‘method of programming’ – associating a (finite) object y with a *program* p and a (finite) object x .

Kolmogorov went on to observe, crucially, that (ibid, pp.299-300):

"[T]he function $K_\phi(y|x)$ need not be effectively computable (generally recursive) even if it is *a fortiori* finite for any x and y ."

Remark 1 *The proof that $K_\phi(y|x)$ is nonconstructive, freely appealing to *tertium non datur*. I consider this an infelicity. But since it is not an existence proof, rectifying the infelicity by a constructive proof may not be essential*

To the best of my knowledge most proofs of the uncomputability of $K_\phi(y|x)$ are based on the unsolvability of the Halting problem for Turing Machines⁸. Shortly after Kolmogorov’s above paper was published, Zvonkin and Levin, [34], p.92, Theorem 1.5, *b*, provided the result and proof that rationalises the basic principle of stochastic complexity providing the computable approximation to the uncomputable $K_\phi(y|x)$. The significant relevant result is:

Theorem 2 *Zvonkin-Levin*

\exists a general recursive function $H(t,x)$, monotonically decreasing in t , s.t :

$$\lim_{t \rightarrow \infty} H(t,x) = K_\phi(y|x) \quad (2)$$

Remark 3 *This result guarantees, the existence of ‘arbitrarily good upper estimates’ for $K_\phi(y|x)$, even although $K_\phi(y|x)$ is uncomputable. I am not sure this is a claim that is constructively substantiable⁹. How can a noncomputable function be approximated? If any one noncomputable function can be approximated uniformly, then by ‘reduction’ it should be possible, for example, to ‘approximate’, say, the Busy Beaver function. I suspect an intelligent and operational interpretation of the Zvonkin-Levin theorem requires a broadening of the notion of ‘approximation’.*

⁸For example in [6], §7.7, pp.162-8. Incidentally, the section on *Models of Computation* (§7.1, pp.146-7), in this book is quite unreliable and strange, to put it mildly. The presentation of the genesis of the Turing Machine and Church’s Thesis are both incorrect to the point of being absurd.

⁹My view on tis further strengthened by some of the remarks in [6], particularly, p.163, where one reads (italics added):

"The shortest program is not computable, although as more and more programs are shown to produce the string, the estimates from above of the Kolmogorov complexity *converge to the true Kolmogorov complexity*, (the problem, of course, is that *one may have found the shortest program and never know that no shorter program exists*).

These remarks border on the metaphysical! How can one approximate to a true value which cannot be known, by definition?

Universality, (approximate) computability, data compression, the eschewing of ‘*truth*’ (in model selection) – these are, in my reading, the four fundamental building blocks of Rissanen’s methodology. They form the methodological building blocks of algorithmic economics, which in earlier writings I called *Computable Economics* (cf. [28]).

3 Re-reading Jevons in the Light of MDL

"Doubtless there is in nature some invariably acting mechanism, such that from some fixed conditions an invariable result always emerges. But we, with our finite minds and short experience, can never penetrate the mystery of these existences We are in the position of *spectators who witness the production* of a complicated machine, but are not allowed to examine its structure. *We learn what does happen and what does appear*, but if we ask for the reason, the answer would involve an infinite depth of mystery."

[13], p.222; italics added.

William Stanley Jevons, a pioneer of neoclassical economics was implacably opposed to the inductive method. His methodological precepts against the inductive method were cogently presented in his monumental treatise on *The Principles of Science* (*ibid*, henceforth referred to as *TPOS*). However, a close reading of its almost 800 pages, against the backdrop of some knowledge of the principles underpinning the MDL principle has convinced me that the Jevonian opposition to the inductive method is untenable. In this section a sketch of my re-interpretation of *TPOS* as a treatise supporting what I have in earlier writings called *The Modern Theory of Induction* ([28], Chapter 5) is outlined.

3.1 Background

"What especially characterised Jevons’s view of logical method was the prominence he attached to the combination of formal and empirical principles *through the inverse application of the theory of probability*"

[14], p.638; italics added

TPOS, a book of almost 800 dense pages refers to almost every known Western natural philosopher without, however, a single mention of *William of Ockham*, *Occam’s Razor* or *Ockham’s Principle*¹⁰! The closest he gets to anything like a (dismissive) mention of *Occam’s Razor* is when he rejects Newton’s *Rule 1*

¹⁰The general literature seems to refer to William of Ockham but Occam’s Razor; hence I retain this schizophrenia in my own spelling. Furthermore, Ockham’s own most often stated version of the principle named after him seems to have been: ‘*Pluralitas non est ponenda sine necessitate*’ – plurality is not be posited without necessity. The more commonly attributed version: ‘*Entia non sunt multiplicanda sine necessitate*’ – entities must not be multiplied without necessity – appears not to have been used by him (cf. [3], p.xxi).

for *Natural Philosophy* in the *Principia* as irrelevant for any inductive purpose, let alone for acting as an anchor to eliminate *inductive indeterminacy*:

"It is by false generalisation, again, that the laws of nature have been supposed to possess that perfection which we attribute to *simple* forms and relations. ... Newton seemed to adopt the questionable axiom that nature always proceeds in the *simplest* way; in stating his first rule of philosophising, he adds: 'To this purpose the philosophers say, that nature does nothing in vain, when less will serve; for nature is pleased with *simplicity*, and affects not the pomp of superfluous causes.'..... *Simplicity* is naturally agreeable to a mind of limited powers, but to an infinite mind all things are simple."

TPOS, p.625; italics added.

Is Jevons suggesting, in the context of his times, beliefs and traditions, that the omnipotence and omniscience of the architect of the laws of nature – the designer of the ‘complicated machine’ – are such that we are as likely to witness the ‘productions of a complicated machine’ as to a simple one¹¹. Jevons may have been trying to make the point that Newton’s was a metaphysical assumption and that we have no grounds for assuming anything about structure in the absence of empirical evidence to the contrary¹². However, Jevons, who was almost as obsessed with consistency as he was with deduction¹³, did not obey his own precepts when it came to choosing the order and degree of equations to fit observed data. In such an example he argues clearly in favour of choosing the *simplest* hypothesis, at least in the first instance:

"It is a *general rule* in quantitative investigation that we commence by discovering linear, and *afterwards* proceed to elliptic or more *complicated* laws of variation."

TPOS, p.474; italics added.

Perhaps, given the times and context, one can be generous to Jevons – more generous than he was to Newton and more, also, than Marshall was to Jevons – and suggest that he was doubtful about any reliance on Occam’s Razor because he did not feel it possible to give a rigorous, invariant, analytical definition of *simplicity*. I think, therefore, it may be reasonable to assume, counterfactually, that Jevons would have accepted the use of Occam’s Razor in hypothesis selection and inductive inference had it been possible to demonstrate that it was

¹¹I cannot but reflect on Einstein’s wise maxim when faced with the Great Scorer’s devise, ‘Subtle is the Lord, but malicious he is not’ – Raffiniert ist der Herrgott aber boshaft ist er nicht – and wish Jevons had shown some humility in the face of Einstein’s undisputed predecessor’s, i.e., Newton’s, own methodological maxims.

¹²The Einsteinian example of the way he reasoned his way towards the general theory of relativity from the special theory is clearly described and discussed by Kemeny, [15]. This example is paradigmatic, of the role of *simplicity* in hypothesis selection and formation, in the logic of scientific practice.

¹³He would, surely, find it uncomfortable to live in a post-Gödelian world where consistency has been dethroned from its crowning place in the deductive enterprise!

possible to define, rigorously, the notion of simplicity. After Solomonoff – not a little influenced by Keynes – and Rissanen, re-reading Jevons and substantiating a rigorous method of inductive inference is not the most difficult task for a philosophy of science. This will be attempted in the last sub-section of this section, after first summarising the Jevonian vision of inductive indeterminacy in the next sub-section.

3.2 The Jevonian Vision on Induction and its Indeterminacy

"Combining insight and error, he spoilt brilliant suggestions by erratic and atrocious arguments. His application of inverse probability to the inductive problem is crude and fallacious, but the idea which underlies it is substantially good. ... There are few books, so superficial in argument yet suggesting so much as Jevons's *Principles of Science*."

Keynes, [17], p.204

I shall summarise, rather telegraphically and in an inelegant numbered-list format, Jevons's precepts on inductive inference. This will, then, enable me to refer to them conveniently in the next section when a simple case is made to encapsulate the Jevonian vision in the modern inductive fold.

The following twelve points summarise, however audacious the task of encapsulating summarily, a sustained criticism of the inductive method, spread over a discursive book of almost 800 pages (all quotes in this list are from *TPOS*):

1. "The theory of inductive inference stated [in *TPOS*] was suggested by the study of the Inverse Method of Probability." (p.265)
2. Induction is the inverse operation of deduction. (p.121)
3. Induction is *perfect* when an enumeration of all possible instances of the phenomenon under consideration is feasible, at least in principle. (pp.146-7)
4. Induction is *imperfect* in case the 'enumeration', as in (3), is infeasible.
5. The results of imperfect induction are, therefore, never more than probable:
 "Only in proportion as our induction approximates to the character of *perfect induction*, does it approximate to certainty. The amount of uncertainty corresponds to the probability that other objects than those examined may exist and *falsify our inferences*; ...". (p.229; italics added)
6. The number of instances of any inductive phenomenon is, at most, denumerably infinite; and the number of alternative hypotheses that may be entertained to account for any given inductive phenomenon is, at most, denumerably infinite.

7. Inductive processes are those, and only those, that generate general laws such that the hypothesis underlying them ‘yield deductive results in accordance with experience.’
8. "That process only can be called induction which gives general laws, and it is by the subsequent employment of deduction that we anticipate particular events. I hold that in all cases of inductive inference we must invent hypotheses, until we fall upon some hypotheses which yields deductive results in accordance with experience." (p.226-8)
9. The extraction of general laws, from a denumerably infinite set of plausible hypotheses, proceeds by way of applying the ‘inverse method of probability’ (i.e., using Bayes’s Rule):

"[I]n all cases ... of inductive inference where we seem to pass from some particular instances to a new instance, we ... form an hypothesis as to the logical conditions under which the given instances might occur; we calculate inversely the probability of that hypothesis and compounding this with the probability that a new instance would proceed from the same conditions, we gain the absolute probability of occurrence of the new instance in virtue of this hypothesis. But as several, or many, or even an infinite number of mutually inconsistent hypothesis may be possible, we must repeat the calculation for each such conceivable hypothesis, and then the complete probability of the future instances will be the sum of the separate probabilities." (p.268)

This description indicates that Jevons’s inductive method, despite its rhetoric about being simply ‘the inverse of deduction’, is nothing other than a simple Bayesian procedure.
10. However, there is no rule or uniform principle on the basis of which it is possible to assign priors to implement ‘the inverse method of probability’ in the mechanical way in which deductive rules can be applied:

"To assign the antecedent probability of any proposition, may be a matter of difficulty or impossibility, and one with which logic and the theory of probability have little concern." (p.211-2)
11. "All logical inference involves classification [and it] is not really distinct from the process of perfect induction. [But] there will be no royal road to the discovery of *the best system* and it will even be *impossible to lay down the rules of procedure* to assist those who are in search of good arrangement." (pp.673-90; italics added)
12. The Ramean Tree (pp.702-3), is an encapsulation of the exhaustive method of classification.

3.3 Disciplining Jevonian Inductive Indeterminacies in a Post-Solomonoff MDL World

"[T]he *most probable* cause of an event which has happened is that which would *most probably* lead to the even supposing the cause to exist."

TPOS, p.243; italics added.

I claim that ‘*most probable*’, in the above Jevonian sense of being encapsulated within the inverse probability framework, is equivalent to the precise *recursion theoretic inductive inference* concept of *simplest* and it removes, effectively, the much vaunted *indeterminacy of induction*. The fundamental notion of the modern theory or recursion theoretic induction can be stated as the following proposition:

Proposition 4 *An event with the highest probability of occurring is also that which has the simplest description*

Let me give a brief and elementary sketch of the kind of analysis that makes such an equivalence possible – i.e., to be able to use Occam’s Razor to eliminate the indeterminacy in the ‘inverse probability’ method, correctly identified by Jevons. Consider a standard version of Bayes’s rule:

$$P(H_i|E) = \frac{P(E|H_i) P(H_i)}{\sum_i P(E|H_i) P(H_i)} \quad (3)$$

Where, apart from absolutely standard, textbook interpretations of all variables and notations, the only implicit novelty – for a Jevonian vision – is the assumption of a denumerable infinity of hypotheses (i.e., above, §3.2:(6)-(7)). This, in a standard inverse probability exercise, E , the class of ‘observed’ events, and $P(H_i)$ are given; Jevons’s inductive inference problem is, then, to find the ‘most probable’ H_i that would ‘most probably’ lead to the observed event of relevance. To get the perspective I want, rewrite (3) as:

$$-\log P(H_i|E) = -\log P(E|H_i) - \log P(H_i) + \log P(E) \quad (4)$$

where the last term on the r.h.s of (4) is a shorthand expression for the denominator in (3) which, in turn, is the normalising factor in such inverse probability exercises.

Now, finding the Jevonian ‘most probable hypothesis’ is equivalent to determining that H_i , w.r.t which (4) is *minimised*. However, in (4), $\log P(E)$ is invariant w.r.t H_i ; hence the problem is to minimise (w.r.t., H_i):

$$-\log P(E|H_i) - \log P(H_i) \quad (5)$$

However, it is clear that the problem of indeterminacy remains so long as we do not have a principle on the basis of which the *prior* cannot be assigned *universally*.

Recall, now, that the Jevonian inductive enterprise is supposed to interpret a class of observations, events, data, etc., – ‘the production of a complicated machine’ – in terms of a denumerable infinity of hypotheses, in such a way that a general law is formalised from which, by deductive processes, the outcomes with which one began are generated (cf. above, §3.2, (2), (7)-(9)). These entities are formalised – in pre-set theoretic days – in terms of logical and mathematical formulas. As far as the requirements of the logic of the inductive method recommended in *TPOS* is concerned, we need only formalise, at most, a denumerable infinity of outcomes in an observation space, and there is a similar quantitative upper bound for the number of hypotheses. Thus the space of computable numbers is sufficient for this formalisation exercise.

Suppose, now, that every element in the outcome space and every potential hypothesis – being denumerably infinite – is associated with a positive integer, perhaps ordered lexicographically. In *TPOS* every outcome and every hypothesis is framed as a logical proposition. Every such proposition can, therefore, be assigned one of the computable numbers and they, in turn, can be processed, say, by a Turing Machine. Next, the *binary codes* for the assigned computable numbers can be constructed, and thereby they can also be given a precise quantitative measure in terms of their counts in *bits*. Thus the basic result of modern recursion theoretic inductive inference, summarised in the above proposition, results from the following *Rissanen Rule of MDL Inductive Inference*:

Proposition 5 *Rule of Induction*¹⁴

The ‘best theory’ is that which minimizes the sum of:

- (a). The length, in bits, of the number theoretic representation of the denumerable infinity of hypothesis;
- (b). The length, in bits, of the elements of the space of outcomes (also, by assumption, at most, denumerably infinite);

The conceptual justification for this ‘rule’ as the underpinning for Proposition 4 is something like the following reasoning> if the elements of the observation space (*E*) have any *patterns* or *regularities*, then they can be encapsulated in a *law*, on the basis of some *hypothesis*. The idea that the *best law* is that which can extract and summarise the *maximum amount of regularities* or patterns in *E* and represent them *most concisely* captures the workings of Occam’s razor in an inductive exercise. In homely terms: if two hypotheses can encapsulate the patterns in the data, then choose the more concise one.

The final link in this inductive saga is a universal formula for the prior in the inverse probability exercise.

Proposition 6 \exists a probability measure $m(\cdot)$ that is universal (in the sense of being invariant except for an inessential additive constant) such that:

$$\log_2 m(\cdot) \approx K(\cdot) \tag{6}$$

¹⁴The problem of summing an infinite sum has to be resolved by some kind of standard normalization procedure in the case, as here, of denumerable infinity of hypotheses. I shall ignore this detail here.

where, $K(\cdot)$: the Kolmogorov complexity of the best theory generated in the implementation of the rule of induction.

I think this closes the circle consistently with the aims set forth in *TPOS* for an inductive exercise. Thus, I rest my case for Jevons, after Solomonoff-Rissanen, as an inductivist.

4 Complexity of an Undecidable Inference in a Dynamical System

"[T]he question of the decidability of the Mandelbrot set has another justification. It can partly answer and give insight to the question: can one *decide* if a differential equation is chaotic?"

[2], p.5; italics added.

I have had to tackle formal undecidabilities in economic dynamics. One of the formal propositions I have derived in economic dynamics relates to the non-effectivity of policy in a complex dynamic economy. In trying to resolve some dissatisfaction with this result, I have been influenced by some of Rissanen's methodological precepts. An outline of a result, the framework and some conjectures are given in this section.

One of the keys to Rissanen's inference methodology lies in eschewing the search for 'true' models that give rise to observable phenomena which have to be explained. Taking a cue from such a methodology I want to pose the following problem: given the observables of a dynamical system, is it possible to *infer* interesting properties that characterise its basins of attraction? In view of Rice's theorem in classical recursion theory – or, alternatively, due to the ubiquity of the unsolvability of the Halting Problem for Turing Machines – it is often impossible to infer whether observable data is sufficient to decide membership in a set, unless the set is characterised trivially.

Let me first provide the formal background in a general way.

I shall have to assume familiarity with the formal definition of a dynamical system (cf. for example, the obvious and accessible classic, [10] or the more modern, [4]), the necessary associated concepts from dynamical systems theory and all the necessary notions from classical computability theory (for which the reader can, with profit and enjoyment, go to a classic like [26] or, at the frontiers, to [5]). Just for ease of reference the bare bones of relevant definitions for dynamical systems are given below in the usual telegraphic form¹⁵. An intuitive understanding of the definition of a 'basin of attraction' is probably sufficient for a complete comprehension of the result that is of interest here - provided there is reasonable familiarity with the definition and properties of Turing Machines (or

¹⁵In the definition of a dynamical system given below I am not striving to present the most general version. The basic aim is to lead to an intuitive understanding of the definition of a basin of attraction so that the main theorem is made reasonably transparent. Moreover, the definition given below is for scalar ODEs, easily generalizable to the vector case.

partial recursive functions or equivalent formalisms encapsulated by Church's Thesis).

Definition 7 *The Initial Value Problem (IVP) for an Ordinary Differential Equation (ODE) and Flows.* Consider a differential equation:

$$\dot{x} = f(x) \tag{7}$$

where x is an unknown function of $t \in I$ (say, t : time and I an open interval of THE REAL LINE) and f is a given function of x . Then, a function x is a solution of (7) on the OPEN INTERVAL I if:

$$\dot{x}(t) = f(x(t)), \forall t \in I \tag{8}$$

The initial value problem (ivp) for (7) is, then, stated as:

$$\dot{x} = f(x), \quad x(t_0) = x_0 \tag{9}$$

and a solution $x(t)$ for (9) is referred to as a solution through x_0 at t_0 . Denote $x(t)$ and x_0 , respectively, as:

$$\varphi(t, x_0) \equiv x(t), \text{ and } \varphi(0, x_0) \equiv x_0 \tag{10}$$

where $\varphi(t, x_0)$ is called the **flow** of $\dot{x} = f(x)$.

Definition 8 Dynamical System

If f is a C^1 function (i.e., the set of all differentiable functions with continuous first derivatives), then the **flow** $\varphi(t, x_0), \forall t$, induces a **map** of $U \subset \mathbb{R}$ into itself, called a C^1 **dynamical system on \mathbb{R}** :

$$x_0 \longmapsto \varphi(t, x_0) \tag{11}$$

if it satisfies the following (one-parameter group) properties:

1. $\varphi(0, x_0) = x_0$
2. $\varphi(t + s, x_0) = \varphi(t, \varphi(s, x_0)), \forall t \ \& \ s$, whenever both the l.h and r.h side maps are defined;
3. $\forall t, \varphi(t, x_0)$ is a C^1 map with a C^1 inverse given by: $\varphi(-t, x_0)$;

Remark 9 A geometric way to think of the connection between a **flow** and the induced **dynamical system** is to say that the flow of an **ODE** gives rise to a dynamical system on \mathbb{R} .

Remark 10 It is important to remember that the **map** of $U \subset \mathbb{R}$ into itself may **not** be defined on all of \mathbb{R} . In this context, it might be useful to recall the distinction between partial recursive functions and total functions in classical recursion theory.

Definition 11 Invariant set

A set (usually compact) $S \subset U$ is **invariant** under the flow $\varphi(.,.)$ whenever $\forall t \in \mathbb{R}, \varphi(.,.) \subset S$.

Definition 12 Attracting set

A closed invariant set $A \subset U$ is referred to as the **attracting set** of the flow $\varphi(t, x)$ if \exists some neighbourhood V of A , s.t $\forall x \in V \ \& \forall t \geq 0, \varphi(t, x) \in V$ and:

$$\varphi(t, x) \rightarrow A \text{ as } t \rightarrow \infty \quad (12)$$

Remark 13 It is important to remember that in dynamical systems theory contexts the attracting sets are considered the **observable** states of the dynamical system and its flow.

Definition 14 The basin of attraction of the attracting set A of a flow, denoted, say, by Θ_A , is defined to be the following set:

$$\Theta_A = \cup_{t \leq 0} \varphi_t(V) \quad (13)$$

where: $\varphi_t(.)$ denotes the flow $\varphi(.,.), \forall t$.

Remark 15 Intuitively, the basin of attraction of a flow is the **set of initial conditions** that eventually leads to its attracting set - i.e., to its limit set (limit points, limit cycles, strange attractors, etc). Anyone familiar with the definition of a Turing Machine and the famous Halting problem for such machines - or, alternatively, Rice's theorem - would immediately recognise the connection with the definition of basin of attraction and suspect that my main result is obvious.

Definition 16 Dynamical Systems capable of Computation Universality:

A dynamical system capable of computation universality is one whose defining initial conditions can be used to program and simulate the actions of any arbitrary Turing Machine, in particular that of a Universal Turing Machine.

Proposition 17 Dynamical systems characterizable in terms of limit points, limit cycles or 'chaotic' attractors, called 'elementary attractors', are not capable of universal computation.

Proposition 18 Only dynamical systems whose basins of attraction are poised on the boundaries of elementary attractors are capable of universal computation.

Theorem 19 There is no effective procedure to decide whether a given observable trajectory is in the basin of attraction of a dynamical system capable of computation universality

Proof. The first step in the proof is to show that the basin of attraction of a dynamical system capable of universal computation is recursively enumerable but

not recursive. The second step, then, is to apply Rice's theorem to the problem of membership decidability in such a set.

First of all, note that the basin of attraction of a dynamical system capable of universal computation is recursively enumerable. This is so since trajectories belonging to such a dynamical system can be effectively listed simply by trying out, systematically, sets of appropriate initial conditions.

On the other hand, such a basin of attraction is not recursive. For, suppose a basin of attraction of a dynamical system capable of universal computation is recursive. Then, given arbitrary initial conditions, the Turing Machine corresponding to the dynamical system capable of universal computation would be able to answer whether (or not) it will halt at the particular configuration characterising the relevant observed trajectory. This contradicts the unsolvability of the Halting problem for Turing Machines.

Therefore, by Rice's theorem, *there is no effective procedure to decide whether any given arbitrary observed trajectory is in the basin of attraction of such recursively enumerable but not recursive basin of attraction.* ■

Remark 20 *There is a 'monumental' mathematical 'fudge' in my proof of the recursive enumerability of the basin of attraction: how can one try out, 'systematically', the set of uncountable initial conditions lying in the appropriate subset of \mathfrak{R} ? Of course, this cannot be done and the theorem is given just to give an idea of the problem that I want to consider.*

Keeping the framework and the questions in mind, one way to proceed would be to constructivise the basic IVP problem for ODEs and then the theorem can be applied consistently. It will require too much space and time to do so within the scope of this paper. Instead, I shall adopt a slightly devious method.

Consider the following Generalized Shift (GS) map ([19],[20]):

$$\Phi : \wp \rightarrow \sigma^{F(\wp)} [\wp \oplus G(\wp)] \quad (14)$$

Where:

\wp : (bi-infinite) symbol sequence;

F : mapping from a finite subset of \wp to the integers;

G : mapping from a finite subset of \wp into \wp ;

σ : a shift operator;

The given 'finite subset of \wp ', on which F and G operate is called the *domain of dependence (DOD)*.

Let the given symbol sequence be, for example:

$$\wp \equiv \{\dots p_{-1} p p_{=1} \dots\} \quad (15)$$

Then:

$\wp \oplus G(\wp) \Rightarrow$ replace DOD by $G(\wp)$.

$\sigma^{F(\wp)} \Rightarrow$ *shift* the sequence left or right by the amount $F(\wp)$

Remark 21 *In practice, a GS is implemented by denoting a distinct position on the initially given symbol sequence as, say, p_0 and placing a ‘reading head’ over it. It must also be noted that $p_i \in \varnothing, \forall i = 1, 2, \dots$ could, for example, denote whole words from an alphabet, etc., although in practice it will be 0, 1 and \circ (‘dot’). The ‘dot’ will signify that the ‘reading head’ will be placed on the symbol to the right of it.*

The following results about Generalized Shift maps are relevant for my discussion:

Proposition 22 *Any GS is a nonlinear (in fact, piecewise linear) dynamical system capable of universal computation; hence they are universal dynamical systems and are equivalent to some constructible Universal Turing Machine.*

Thus the GS is capable of universal computation and it is minimal in a precisely definable sense (see [19] and [20] for full details). It is also possible to construct, for each such generalized shift dynamical system¹⁶, an equivalent UTM that can simulate its dynamics, for sets of initial conditions. Now consider the observable set of the dynamical system, $y \in A$; given the UTM, say \mathbb{U} , corresponding to φ , the question is: for what set of initial conditions, say x , is y the halting state of \mathbb{U} . Naturally, by the theorem of the unsolvability of the Halting problem, this is an undecidable question. This is the theorem used in demonstrating the uncomputability of $K_\phi(y|x)$. However, by the above Zvonkin-Levin theorem, we know that the *existence* of ‘arbitrarily good upper estimates’ for $K_\phi(y|x)$, even although $K_\phi(y|x)$ is uncomputable.

Now, taking a cue from Rissanen’s methodological point about the irrelevance of ‘true’ models, but only of models that can explain the data minimally, let me consider the above (minimal) universal dynamical system as canonical for any question about membership in attracting sets, A . What is the complexity of $K_{\mathbb{U}}(p|x)$? By definition it should be:

$$K_{\mathbb{U}}(y|x) = \left\{ \begin{array}{l} \min_{\mathbb{U}(p,x)=y} l(p) \\ \infty \nexists p \text{ s.t } \phi(p, x) = y \end{array} \right.$$

The meaning, of course, is: the minimum over all programs, p , implemented on \mathbb{U} , with the given initial condition, x , which will stop at the halting configuration, y . The above theorem formalizes the notion that there is no general algorithmic procedure to decide any such membership.

Remark 23 *Why is it important to show the existence of the minimal program? Because, is the observed y corresponds to the minimal program of the dynamical system, i.e., of \mathbb{U} , then it is capable of computation universality; if there is no minimal program, the dynamical system is not interesting! A monotone decreasing set of programs that can be shown to converge to the minimal program is analogous to a series of increasingly complex finite automata converging to a*

¹⁶They can also encapsulate smooth dynamical systems in a precise sense. I have described the procedure, summarising a part of Chris Moore’s approach, in [28], Chapter 4.

TM. What we have to show is that there are programs converging to the minimal program from above and below, to the border between two basins of attractions.

Thus, behind every undecidable proposition – at least in principle – there is an inference principle which may or may not suggest an ‘approximation’ strategy to ‘decide the undecidable’. After all, Gödel himself thought that the undecidable may become decidable by going ‘upwards’, so to speak, in strengthening the axiom systems; surely, there must be a practical way of going in the opposite direction to locate the borders of the decidable as approximation to the undecidable, too. Naturally, I expect these highly speculative conjectures to apply, *pari passu*, to the computable-uncomputable divide, too.

5 Concluding Thoughts

"Inductive processes have formed, of course, at all times a vital, habitual part of the mind's machinery. Whenever we learn by experience, we are using them. But in the logic of the schools they have taken their proper place slowly."

John Maynard Keynes, [16], p.241.

It is to Jorma Rissanen's lasting credit that he has, almost single-handedly developed a scientific method to make this ‘habitual part of the mind's machinery’ entirely and rigorously algorithmic. Thus, he belongs to the modern scientific movement towards an algorithmic approach to statistics, randomness and information. As an economist, I have strived to develop an analogous field of algorithmic economics, where stochastic complexity and the MDL principle are as central as algorithmic randomness, computability theory and computational complexity theory. Learning and induction – indeed, learning as induction – is a central topic at the frontiers of economics. The frontier researchers remain blissfully ignorant of the algorithmic approach to learning and inductive inference, randomness and information. This is strange in a subject which prides itself on placing the role of scarce information and its husbanding in its citadel.

Economics is singularly free of an algorithmic vision. The mathematics of economic theory is dominated by Bourbakian thinking. The methodology of statistical inference in economics is equally stone-aged.

The success of Jorma Rissanen's single-handed, even single-minded, efforts to inject a new algorithmic vision into statistical methodology, particularly in inference, estimation and prediction theories, is heartening for those of us who find ourselves at the fringes of mathematical economics in view of our algorithmic vision.

I believe Jorma Rissanen's work contributes a missing link to the great Finnish tradition of work in inductive logic, one which was most cogently stated by Hilary Putnam in that period of an interregnum between the growing systematization of the philosophy of inductive logic and the emergence of the recursion

theoretic inductive movement¹⁷ ([22], p.297):

"[W]e may think of a system of inductive logic as a design for a 'learning machine': that is to say, a design for a computing machine that can extrapolate certain kinds of empirical regularities from the data with which it is supplied."

Jorma Rissanen, together with Ray Solomonoff, have pioneered and 'patented' not only the design for a 'learning machine'; they have actually built it.

¹⁷In particular, it must be remembered that Solomonoff's work straddles the two traditions and his two path-breaking contributions appeared almost before the proverbial ink was dry on Putnam's seminal contribution.

References

- [1] Barron, A. R, Jorma Rissanen & B. Yu, (1998), *The Minimum Description Length Principle in Coding and Modelling*, **IEEE Transactions on Information Theory**, Vol. 44, #6, pp. 2743-60.
- [2] Blum, Lenore, Felipe Cucker, Michael Shub and Steve Smale (1998), **Complexity and Real Computation**, Springer Verlag, New York.
- [3] Boehner, P, (1990), **Okham – Philosophical Writings**, Hackett, Indianapolis.
- [4] Brin, Michael and Garrett Stuck (2002): **Introduction to Dynamical Systems**, Cambridge University Press, Cambridge.
- [5] Cooper, S Barry (2004): **Computability Theory**, Chapman & Hall/CRC, Boca Raton and London
- [6] Cover, Thomas. M & Joy A. Thomas, (1991), **Elements of Information Theory**, John Wiley & Sons, Inc., New York & Chichester.
- [7] Dordick, Rowan, (1988), *Understanding the ‘go’ of it*, **IBM Research Magazine**, Winter.
- [8] Gács, Péter, John T. Tromp and Paul M. B. Vitányi, (2001), *Algorithmic Statistics*, **IEEE Transactions on Information Theory**, Vol.47, #6, September, pp.2443-63
- [9] Hilpinen, Risto, (1968), **Rules of Acceptance and Inductive Logic**, Acta Philosophica Fennica, 22, North-Holland, Amsterdam.
- [10] Hirsch, Morris. W, Stephen Smale & Robert L. Devaney, (2004), **Differential Equations, Dynamical Systems & An Introduction to Chaos**, Academic Press, an Imprint of Elsevier, San Diego & London.
- [11] Hintikka, Jaakko & Ilkka Niiniluoto, (1966), *A Two-Dimensional Continuum of Inductive Methods*, pp. 113-32, in: **Aspects of Inductive Logic**, edited by Jaako Hintikka & Patrick Suppes, North-Holland, Amsterdam.
- [12] Hintikka, Jaakko & Ilkka Niiniluoto, (1980), *An Axiomatic Foundation for the Logic of Inductive Generalization*, Chapter 7, pp.157-81, in: **Studies in Inductive Logic and Probability, Volume II**, edited by Richard C. Jeffrey, University of California Press, Berkeley and Los Angeles.
- [13] Jevons, W. Stanley, (1877), **The Principles of Science: A Treatise on Logic and Scientific Method**, Second Edition, Revised, Macmillan & Co., London & New York.
- [14] Johnson, W. E, (1895), *Logic and Political Economy*, pp.637-8, in: **Dictionary of Political Economy**, Volume 2, edited by R.H. Inglis Palgrave, Macmillan, London.

- [15] Kemeny, J.G, (1953), *The Use of Simplicity in Induction*, **The Philosophical review**, Vol.62, pp.391-408.
- [16] Keynes, John Maynard, (1921), **A Treatise on Probability**, Reprinted as: Volume VIII, *The Collected Writings of John Maynard Keynes*, Macmillan, Cambridge University Press, for the Royal Economic Society, London.
- [17] Keynes, John Maynard, (1973, [1936]), *William Stanley Jevons*, Reprinted in: Volume X, **The Collected Writings of John Maynard Keynes**, Macmillan, Cambridge University Press, for the Royal Economic Society, London.
- [18] Kolmogorov, A.N, (1968), *Three Approaches to the Definition of the Concept of the "Amount of Information"*, pp. 293-302, **Selected Translations in Mathematical Statistics and Probability**, Volume 7, American Mathematical Society, Providence, Rhode Island.
- [19] Moore, Christopher, (1990), *Unpredictability and Undecidability in Dynamical Systems*, **Physical Review Letters**, Vol.64, #4, 14 May, pp.2354-7.
- [20] Moore, Christopher, (1991), *Generalized Shifts: Unpredictability and Undecidability in Dynamical Systems*, **Nonlinearity**, Vol.4, pp. 199-230.
- [21] Niiniluoto, Ilkka & Raimo Tuomela, (1973), **Theoretical Concepts and Hypothetico-Inductive Inference**, Dordrecht, Boston.
- [22] Putnam, Hilary, (1975 [1963]), *Probability and Confirmation*, Reprinted in: **Mathematics, Matter and method – Philosophical Papers**, Vol. 1, Cambridge University Press, Cambridge.
- [23] Rissanen, Jorma, (1978), *Modelling by Shortest Data Description*, **Automatica**, Vol.14, # 5, September, pp.465-71.
- [24] Rissanen, Jorma, (1986), *Stochastic Complexity and Modeling*, **The Annals of Statistics**, Vol.14, #3, pp.1080-1100.
- [25] Rissanen, Jorma, (1987), *Stochastic Complexity*, **Journal of the Royal Statistical Society, Series B**, Vol.49, #3, pp.223-39 & pp.252-65.
- [26] Rogers, Hartley, Jr., (1967): **Theory of Recursive Functions and Effective Computability**, McGraw-Hill, New York.
- [27] Rustem, Berc, Kumaraswamy Velupillai & John H. Westcott, (1978), *Respecifying the Weighting Matrix of a Quadratic Objective Function*, **Automatica**, Vol.14,# 6, November, pp. 567-82.
- [28] Velupillai, Kumaraswamy, (2000), **Computable Economics**, Oxford University Press, Oxford.

- [29] Velupillai, K. Vela, (2007), *The Impossibility of an Effective Theory of Policy in a Complex Economy*, in: **Complexity Hints for Policy**, edited by David Colander and Massimo Salzano, Springer-Verlag, Berlin and New York.
- [30] Velupillai, K. Vela, (2007), *Demystifying Induction and Falsification: Trans-Popperian Suggestions*, in: **Induction, Metaphysics and Economic Methodology: Reflection on Popperian Themes** edited by Tom Boylan and Pascal O’Gorman, Routledge, London.
- [31] Vitanyi, Paul M.B & Ming Li, (2000), *Minimum Description length Induction, Bayesianism, and Kolmogorov Complexity*, **IEEE Transactions on Information Theory**, Vol.46, # 2, March, pp.446-64.
- [32] von Wright, Georg Henrik, (1951), **A Treatise on Induction and Probability**, Routledge & Kegan Paul, London.
- [33] von Wright, Georg Henrik, (1957), **The Logical Problem of Induction**, Basil Blackwell, Oxford.
- [34] Zvonkin, A.K & L.A. Levin, (), *The Complexity of Finite Objects and the Development of the Concepts of Information and Randomness by Means of the Theory of Algorithms*, **Russian Mathematical Surveys**, Vol.25, # 6, pp. 83-124.