

THE LIMITS OF PREDICTABILITY: INDETERMINISM AND UNDECIDABILITY
IN CLASSICAL AND QUANTUM PHYSICS

by

ALEXANDRE V. KOROLEV

B.Sc., Novosibirsk State University, 1994
M.A. (High Hon.), Novosibirsk State University, 1996

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENT FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Philosophy)

THE UNIVERSITY OF BRITISH COLUMBIA

October 2007

© Alexandre V. Korolev, 2007

ABSTRACT

This thesis is a collection of three case studies, investigating various sources of indeterminism and undecidability as they bear upon in principle unpredictability of the behaviour of mechanistic systems in both classical and quantum physics.

I begin by examining the sources of indeterminism and acausality in classical physics. Here I discuss the physical significance of an often overlooked and yet important Lipschitz condition, the violation of which underlies the existence of anomalous non-trivial solutions in the Norton-type indeterministic systems. I argue that the singularity arising from the violation of the Lipschitz condition in the systems considered appears to be so fragile as to be easily destroyed by slightly relaxing certain (infinite) idealizations required by these models. In particular, I show that the idealization of an absolutely nondeformable, or infinitely rigid, dome appears to be an essential assumption for anomalous motion to begin; any slightest elastic deformations of the dome due to finite rigidity of the dome destroy the shape of the dome required for indeterminism to obtain. I also consider several modifications of the original Norton's example and show that indeterminism in these cases, too, critically depends on the nature of certain idealizations pertaining to elastic properties of the bodies in these models. As a result, I argue that indeterminism of the Norton-type Lipschitz-indeterministic systems should rather be viewed as an artefact of certain (infinite) idealizations essential for the models, depriving the examples of much of their intended metaphysical import, as, for example, in Norton's antifundamentalist programme.

Second, I examine the predictive computational limitations of a classical Laplace's demon. I demonstrate that in situations of self-fulfilling prognoses the class of

undecidable propositions about certain future events, in general, is not empty; any Laplace's demon having *all* the information about the world now will be unable to predict all the future. In order to answer certain questions about the future it needs to resort occasionally to, or to consult with, a demon of a higher order in the computational hierarchy whose computational powers are beyond that of any Turing machine. In computer science such power is attributed to a theoretical device called an *Oracle* – a device capable of looking through an infinite domain in a finite time. I also discuss the distinction between *ontological* and *epistemological* views of determinism, and how adopting Wheeler-Landauer view of physical laws can entangle these aspects on a more fundamental level.

Thirdly, I examine a recent proposal from the area of quantum computation purporting to utilize peculiarities of quantum reality to perform *hypercomputation*. While the current view is that quantum algorithms (such as Shor's) lead to re-description of the complexity space for computational problems, recently it has been argued (by Kieu) that certain novel quantum adiabatic algorithms may even require reconsideration of the whole notion of computability, by being able to break the Turing limit and "compute the non-computable". If implemented, such algorithms could serve as a physical realization of an Oracle needed for a Laplacian demon to accomplish its job. I critically review this latter proposal by exposing the weaknesses of Kieu's quantum adiabatic demon, pointing out its failure to deliver the purported hypercomputation. Regardless of whether the class of hypercomputers is non-empty, Kieu's proposed algorithm is not a member of this distinguished club, and a quantum computer powered Laplace's demon can do no more than its ordinary classical counterpart.

TABLE OF CONTENTS

Abstract	i
Table of Contents	iii
List of Figures	vii
Acknowledgments	ix
Dedication	xi
 1 Introduction	 1
 2 Part I: Indeterminism, Asymptotic Reasoning and Physically	
Inadmissible Idealizations in Classical Physics	10
2.1 Introduction	11
2.2 The Mass on the Dome	14
2.2.1 Setting the Stage: The Original Formulation	14
2.2.2 The Mass on the Dome: Cartesian Perspective	17
2.2.3 The Mass on the Dome vs. the Mass on the Pinnacle	19
2.2.4 The Mass on the Dome or the Mass in the Air?	22
2.2.5 The Mass on the Dome, Modified	26
2.3 The Lipschitz Condition	27
2.4 Elastic Deformations and Physically Inadmissible Idealizations	34
2.4.1 Models and Idealizations	34
2.4.2 Elastic Deformations and Idealization of a Concentrated Force ..	35
2.4.3 Modification 1: The Dome with a Pinnacle on Top	41
2.4.4 Modification 2: The Rope-Sliding-the-Edge Example	42

2.4.5	Bending of Rods	45
2.4.6	Modification 3: The Rope on the Spherical Top Dome	48
2.5	Asymptotic Reasoning in Philosophy of Science	52
2.5.1	Introduction	52
2.5.2	The Fallacy of Infinite Reasoning	52
2.5.3	Stability and Parameter Sensitivity	54
2.5.4	The Lipschitz-Indeterministic Systems and Physically Inadmissible Idealizations	61
2.6	Lipschitz-Indeterministic Solutions and the Markov Condition	63
2.6.1	Generalized Flows in Hydrodynamics	63
2.6.2	Non-Lipschitz Velocity Fields, Regularizations, and the Markov Condition	66
3	Part II: Undecidability and Unpredictability in Mathematics and Physics	71
3.1	Introduction	72
3.2	Undecidability in Formal Logic and Mathematics	74
3.2.1	Incompleteness in Formal Logic	74
3.2.2	Undefinability of Truth	78
3.2.3	Universality and Self-Reference	80
3.2.4	Diagonalization and Reduction	82
3.3	Classical Computation	83
3.3.1	Turing Machines	84
3.3.2	Computability, Recursiveness, Decidability, and	

Semidecidability	87
3.3.3 The Church -Turing Thesis and Universal Turing Machines	88
3.3.4 Classical Complexity Class P	91
3.3.5 Nondeterministic Turing Machines, Class NP , and NP - completeness	92
3.3.6 Probabilistic Turing Machines and Class BPP	94
3.3.7 Many-Faced Undecidability of the Halting Problem	96
3.4 Diophantine Equations and Hilbert's Tenth Problem	101
3.4.1 Introduction	101
3.4.2 Hilbert's Tenth Problem as a Decision Problem	104
3.4.3 Systems of Diophantine Equations	105
3.4.4 Solutions in Integers, Natural Numbers, and Rational Numbers ..	106
3.4.5 Families of Diophantine Equations and Diophantine Sets	109
3.4.6 Undecidability of Hilbert's Tenth Problem	111
3.4.7 Diophantine Machines	116
3.5 Self-Fulfilling Prognoses and Laplacian Demon	118
3.5.1 Self-Reference and Self-Fulfilling Prognoses	118
3.5.2 Undecidability of Self-Fulfilling Prognoses	121
3.6 Undecidability in Physics	130
3.6.1 Introduction	130
3.6.2 The Problem of Forecast of a Mechanistic System	133
3.6.3 Weakly Chaotic Systems and Prediction by Simulation	138
3.6.4 Undecidability and Randomness	147

3.6.5	Modeling Classical and Quantum Measurements	155
4.	Part III: Quantum Hypercomputation and Oracle Machines	157
4.1	Introduction	158
4.2	Quantum Adiabatic Computation	159
4.2.1	Introduction	159
4.2.2	The Class QBP and the Powers of Quantum Computers	161
4.2.3	The Quantum Adiabatic Evolution Computer	165
4.2.3.1	Introduction	165
4.2.3.2	The Satisfiability Problem	166
4.2.3.3	The Quantum Adiabatic Theorem	168
4.2.3.4	The Problem Hamiltonian H_P	171
4.2.3.5	The Initial Hamiltonian H_I	172
4.2.3.6	Adiabatic Evolution	173
4.3	Quantum Adiabatic Hypercomputation	175
4.3.1	Beyond Undecidability: Oracle Machines	175
4.3.2	Oracles and Hypercomputation	180
4.3.3	The Quantum Adiabatic "Hypercomputer"	182
4.3.3.1	Introduction	182
4.3.3.2	How Slowly is Slowly Enough?	185
4.3.3.3	The Same Old Story (Told Quantum Mechanically)	190
5.	Conclusion	193
6.	Bibliography	195

LIST OF FIGURES

Figure 1	Mass sliding on the dome	14
Figure 2	Mass on the pinnacle	20
Figure 3	Mass on the dome (proper)	21
Figure 4	Passing over the dome apex	23
Figure 5	Mass detachment in the vicinity of the rim	24
Figure 6	No mass detachment in the vicinity of the apex	25
Figure 7	The direction field for $\alpha = 3/4$	29
Figure 8	Fig. 8	30
Figure 9	Fig. 9	31
Figure 10	Deformation of the dome under the action of a concentrated force	37
Figure 11	Logarithmic infinite well at the apex	40
Figure 12	The dome with a pinnacle on top	41
Figure 13	The rope sliding down the edge	43
Figure 14	The bending of the rope in the immediate vicinity of the edge	45
Figure 15	The bending of the rope in the immediate vicinity of the apex	47
Figure 16	The radius of the curvature of a plain curve	48
Figure 17	The rope sliding down the spherical top dome	49
Figure 18	The "proof" that $2 = \sqrt{2}$	53

Figure 19	Phase-space of a simple gravity pendulum	56
Figure 20	The set of all solutions of (25) for $\alpha = 3/4$ and $g(t) = \cos(t)$	68
Figure 21	The set of all solutions of (26)	69
Figure 22	Nondeterministic Diophantine machine	117
Figure 23	The Bernoulli shift map	143
Figure 24	The suspected relationship of BQP to other problem classes	162

ACKNOWLEDGEMENTS

I gratefully acknowledge the assistance of the members of my supervisory committee – Dr. Steven Savitt, Dr. Paul Bartha, and Dr. Andrew Irvine, of the UBC Philosophy Department, and my internal examiner – William Unruh of the UBC Physics Department. Their guidance, support, comments, criticism and suggestions have made this dissertation possible and helped considerably refine, sharpen, and organize the ideas and arguments contained in it.

My mille grazie and toda rabas go to a long-time friend and colleague of mine, Amit Hagar. I still remember the lounge room of the UBC Philosophy Department where, four years ago, we came up with our first arguments against quantum hypercomputation that laid the foundation for our collaboration and which also appear in Part III of this dissertation.

I am also grateful to Dr. Klementy F. Samohvalov of the Sobolev Institute of Mathematics, whose work has always been a source of inspiration and insights. His analysis of provocative prognoses that he presented on a philosophy seminar at Novosibirsk State University years ago made possible the results on undecidability of the future contingent truths which appear in Part II of this thesis.

I would also like to thank Alexander Roitershtein and Alexei Cheviakov of the UBC Department of Mathematics for useful discussions and suggestions on Norton's mass-on-the-dome example that appears in Part I of the dissertation.

My sincere appreciation must also go to Dr. John Woods, who has made himself familiar with much of my work and generously provided me with much support and encouragement throughout these years.

I should also take this opportunity to thank Oliver Schulte of Simon Fraser University, who has contributed greatly to my graduate studies in North America, and who has taught me a great deal of formal logic and the theory of computation.

I also acknowledge my debt of gratitude to the UBC Philosophy Department that provided me with moral, intellectual, and financial support throughout all the years of my stay in beautiful British Columbia.

Finally, I owe much to my wonderful wife Elena, who has been by my side from the very beginning, through all these years and all the new places that we had to go.

DEDICATION

To my wonderful wife Elena and our children Vladik and Seva

It's hard to predict, especially the future.

Attributed to Niels Bohr¹

1. Introduction

People yearn to know what the future holds. The decisions people make are often based on their present beliefs or expectations about the future, thus the importance of prediction. However, the ability to predict the future is hampered by various obstacles, both of practical and principle nature. Rather than having to deal with the contingencies of human existence, it has been a long tradition in philosophy to introduce special mental constructions called *demons*. Demons as conceptual computational devices often appearing in thought experiments are endowed with specific computational powers that are supposed to enable them to solve specific computational tasks. In this thesis I intend to investigate in principle constraints on predictability for a particular version of Laplacian determinism, as it appears, for example, in Stone (1989) and Svozil (1993). It is argued that, in the view of these constraints – both of ontological (as in Part I) and epistemological (as in Part II) nature – even if classical physics were true, the detailed determination of the future would still be out of human reach.

Historically, the notions of determinism, causality and predictability have been often seen as closely related, if not outright identical. Laplace's famous definition of determinism involves all the three notions inseparably entangled – it starts with a causal characterization and ends with identification of determinism with predictability:

¹ Some believe it to be an old Danish proverb popularized by Niels Bohr (e.g., Kac 1975).

We ought to regard the present state of the universe as the effect of its antecedent state and as the cause of the state that is to follow. An intelligence knowing all the forces acting in nature at a given instant, as well as the momentary positions of all things in the universe, would be able to comprehend in one single formula the motions of the largest bodies as well as the lightest atoms in the world, provided that its intellect were sufficiently powerful to subject all data to analysis; to it nothing would be uncertain, the future as well as the past would be present to its eyes.²

Nowadays, the philosophical literature discussing the interrelations between determinism, causality and predictability is enormous and extremely diverse (e.g., Popper 1950, Russell 1953, Feigl 1953, Bunge 1967, Boyd 1972, Earman 1986 and 2004, Hunt 1987, Stone 1989, Van Kampen 1991, Batterman 1993, Kellert 1993, Bricmont 1995, Batterman and White 1996, Schurz 1996, Schmidt 1998, Hoefer 2004, Bishop 2002, 2003, 2004). Some have maintained that determinism implies predictability while others have maintained that predictability implies determinism. Many have maintained that there are no implication relations between determinism and predictability whatsoever. Some have assumed that in a world of deterministic laws, causality must reign supreme, and some have assumed that in any world where causality is strong enough, determinism must hold. Others have argued, instead, that these two notions are incompatible, and, in any deterministic world complex enough to resemble ours, there is no room for genuine causality.

² Laplace (1820), Preface; translation from Nagel (1961), pp. 281-282.

A great deal of the controversy persists due to lack of a unified consensus of what is understood by the key ingredients that enter the notions of determinism and causality, and there has been a long-standing disagreement of how these concepts should be properly defined. Sobel (1998), to take an example, identifies at least ninety (!) varieties of what the term "determinism" could mean in different contexts. In this thesis I will focus on a particular version of Laplacian determinism, as it appears, for example, in Stone (1989) and Svozil (1993).

Determinism as a *metaphysical* doctrine about the world should be separated from *scientific determinism* – the determinism studied in physical theories (Bishop 2004). As the study of what properties a scientific theory or model must possess in order to be deterministic/indeterministic, scientific determinism is usually much easier to approach in case of concrete theories than investigating the features of scientific or metaphysical determinism in general. Laplacian original characterization as cited above is a paradigm example of an attempt to define scientific determinism. Mark Stone gave a particularly clear characterization of Laplacian determinism specified in the form of jointly necessary and sufficient conditions for determinism in classical particle mechanics, as follows (Stone 1989, see also Bishop 2002, 2003, 2004).

We assume that the physical state of a system is characterized by the values of the positions and momenta of all particles constituting the system at some fixed time t . Furthermore, we assume that a physical state of the system corresponds to a point in state space in an appropriate model (invoking certain idealizations or model assumptions) that allows the description of the system through these values. We then can develop deterministic mathematical models for the evolution of these points in state

space. The following three features have been proposed by Stone as expressing Laplace's vision of determinism (Stone 1989, Kellert 1993, Bishop 2002, 2003, 2004):

(DD) *Differential Dynamics*: there exists an algorithm relating a state of a system at a given time to a state at any other time and the algorithm is not probabilistic.³

(UE) *Unique Evolution*: A given state is always followed (preceded) by the same history of state transitions.⁴

(VD) *Value Determinateness*: Any state of a system can be described with arbitrary small (no-zero) error.⁵

Differential dynamics is motivated by actually existing physical theories that are typically expressed in terms of mathematical equations. These equations, as well as all the initial and boundary conditions, are required to contain no intrinsically probabilistic (as present, for example, in some versions of quantum mechanics) elements in them. Such equations describe the individual trajectories of the points in state space representing states of the system.

The unique evolution requirement is closely associated with DD and expresses the Laplacian belief that systems in classical particle mechanics will repeat their behaviour exactly if the initial and boundary conditions are uniquely fixed and specified.

³ Apart from differential equations, "differential dynamics" can be expressed in the form of difference, integral and integro-differential equations among other possibilities arising in descriptions of physical theories.

⁴ As formulated, UE expresses the uniqueness of state transitions in both temporal directions. It can be easily reformulated to allow for the uniqueness of unidirectional state transitions, resulting, correspondingly, in *futuristic* determinism, and *retrospective* determinism (Earman 1986).

⁵ These descriptions can reflect ontological as well as epistemological constraints of the theory.

Value determinateness is motivated by the Laplacian belief that there is nothing that *in principle* prevents mathematical descriptions of real physical systems with arbitrary accuracy (at least in classical particle mechanics). For example, the ordinary models of classical particle mechanics presuppose precise, definite values for the constants and variables that enter the equations of motions.⁶ It is only with the advent of quantum mechanics the applicability of definiteness to all of physics was questioned.

An alternative characterization of Stone's conditions, due to Svozil (1993), specifies the conditions for determinism in exact recursive-theoretic terms. It assumes that a *recursive evolution function* is "at the heart" of the classical notion of "determinism". Then *strong determinism* or *mechanistic determinism* or simply *mechanism* is taken to be a synonym for recursive-theoretic *total computability*, or *computability in all aspects*. This should not be confused with the above requirement of a recursive evolution function. It is a non-trivial substitution as it requires *all* theoretical entities of the theory to be effectively computable.

For example, since the evolution functions, as well as the initial values and the solutions, are usually defined on continua, such as \mathbf{R} , and since "almost all" (all except the set of measure zero) elements of the continuum are uncomputable, the assumption of an exact (i.e., effectively computable) description of the initial value(s) becomes unrealistic. Here one may wish to resort to the notion of "arbitrarily but finite accuracy parameter values," or "finitely computable parameter values." Since finite numbers are recursively enumerable, this would restore effective computability.

⁶ Clark Glymour takes the VD requirement as a necessary criterion for determinism and cites Peirce and Reichenbach as examples of philosophers who have included this criterion in their analyses of determinism (Glymour 1971, pp. 744-745).

Therefore, in what follows by the term a "deterministic system" we shall understand one having an *effectively computable evolution function*, whereas a system which is totally computable in all of its aspects shall be called "mechanistic". This would constitute our working translation of a talk of Laplace's demon into the language of recursion theory:

A **deterministic theory** has an evolution function which is effectively computable / recursive.

A **mechanistic (strongly deterministic) theory** is effectively computable / recursive in total, i.e., all theoretical entities of the system are effectively computable / recursive. In particular, all initial values, laws and solutions of a mechanistic theory are *recursively enumerable*.⁷

The definitions given above do not directly refer to predicates such as "well defined" or "observable". Also, the question remains open as to whether it is possible to obtain physically meaningful (i.e., compatible with presently existing physical theories) *non-recursive* solutions of recursive initial values and recursive evolution functions. As has been shown by Kreizel (1974) and Pour-El and Richards (1981), such solutions do exist. In order to avoid them (if one wishes to do so), it is necessary to impose further restrictions on physical solutions. Such restrictions, if imposed, should be ultimately motivated by physical considerations, as I do in the Part I introducing the Lipschitz condition to eliminate singularities and divergences in physically meaningful parameters.

⁷ The term "mechanistic" is due to Kreizel (1974).

Further remarks on the recursive-theoretic definition of (Laplacian) determinism are in order:

- (1) Physical determinism in the above defined sense does *not* imply the initial and/or the solution (i.e., the final state) can be represented by an effective computation.
- (2) Computability of the equation of motion *and* the initial value(s) does *not* guarantee computability of the solution, as happens when the solution is non-unique (Kreizel 1974), or obtained by unbounded linear operators, or is a weak solution (Pour-El and Richards 1981). (For *mechanistic* theories uncomputable solutions from computable initial values and computable equations of motions are excluded by definition.)
- (3) Since, from the point of view of coding and information theory, the distinction between the "evolution" and the "initial value" (i.e., some algorithm and its output) seems rather arbitrary, the above distinction between "determinism" and "mechanism" is rather arbitrary as well, and it would probably be more precise to talk of the distinction between non-recursive and recursive (computable and non-computable) ("mechanistic") theories.

This particular version of Laplacian deterministic systems provides a unified recursion theoretic framework in which the issues of determinism / indeterminism, undecidability and computational limitations of predictive conceptual devices – demons – can be studied. Correspondingly, this thesis can be seen as the collection of three independent case studies, investigating the sources of indeterminism in classical physics, computational limitations of classical Laplacian demon as well computational powers of a quantum-computer powered Laplacian demon.

I begin by examining the sources of indeterminism and acausality in classical physics. Here I discuss the physical significance of an often overlooked and yet important Lipschitz condition, the violation of which underlies the existence of anomalous non-trivial solutions in the Norton-type indeterministic systems. I argue that the singularity arising from the violation of the Lipschitz condition in the systems considered appears to be so fragile as to be easily destroyed by slightly relaxing certain (infinite) idealizations required by these models. In particular, I show that the idealization of an absolutely nondeformable, or infinitely rigid, dome appears to be an essential assumption for anomalous motion to begin; any slightest elastic deformations of the dome due to finite rigidity of the dome destroy the shape of the dome required for indeterminism to obtain. I also consider several modifications of the original Norton's example and show that indeterminism in these cases, too, critically depends on the nature of certain idealizations pertaining to elastic properties of the bodies in these models. As a result, I argue that indeterminism of the Norton-type Lipschitz-indeterministic systems should rather be viewed as an artefact of certain (infinite) idealizations essential for the models, depriving the examples of much of their intended metaphysical import, as, for example, in Norton's antifundamentalist programme.

Secondly, I examine the predictive computational limitations of a classical Laplace's demon in situations where prognoses put forth about the future state of a (mechanistic) system actively *provoke* the very events the prognoses are about. Of special interest is a subclass of all such prognoses, so called *self-fulfilling prognoses*, where the very fact of formulating, or putting forth, a prognosis about the state of the system at a certain time in future *initiates*, or *triggers*, a series of changes within the

system, in such a way that at that future moment the system assumes exactly the state described in the prognosis. I demonstrate that in situations of self-fulfilling prognoses the class of undecidable propositions about certain future events is not empty; any Laplace's demon having *all* the information about the world now will be unable to predict all the future. In order to answer certain questions about the future it needs to resort occasionally to, or to consult with, a demon of a higher order in the computational hierarchy whose computational powers are beyond that of any Turing machine. In computer science such power is attributed to a theoretical device called an *Oracle* – a device capable of looking through an infinite domain in a finite time.

Thirdly, I examine a recent proposal from the area of quantum computation purporting to utilize peculiarities of quantum reality to perform supertasks. While the current view is that quantum algorithms (such as Shor's) lead to re-description of the complexity space for computational problems, recently it has been argued (by Kieu) that certain novel quantum adiabatic algorithms may even require reconsideration of the whole notion of computability itself, by being able to "compute the non-computable". If implemented, such algorithms could serve as a physical realization of an Oracle needed for a Laplacian demon to accomplish its job. I critically review this latter proposal by exposing the weaknesses of Kieu's quantum adiabatic demon, pointing out its failure to deliver the purported hypercomputation. Regardless of whether the class of hypercomputers is non-empty, Kieu's proposed algorithm is not a member of this distinguished club, and, when it comes to computing the non-computable, a quantum computer powered Laplace's demon can do no more than its ordinary classical counterpart.

PART I

Indeterminism, Asymptotic Reasoning, and Physically Inadmissible Idealizations in Classical Physics

2.1 Introduction

Abstruse theories like quantum mechanics and general relativity routinely violate common intuitions about causality and determinism. In contrast, classical physics is often assumed to be a paradigm example of a fully deterministic physical theory that never violates these intuitions, or that violates them only in the most extreme circumstances which render such situations as plainly unphysical. A number of authors have argued that this is not so, and that classical physics is a poor choice of hunting ground for such beliefs. A definitive guide to the discussion is John Earman's *A Primer on Determinism* (1986) that collects and discusses various situations that threaten uniqueness of solutions for common differential equations governing dynamics of ordinary classical systems. A more recent attempt by John Norton (2003) presents another simple Newtonian system that seems to exhibit anomalous acausal behaviour in that it allows generation of spontaneous motion of a mass without any external intervention or any change in the physical environment. The latter system is of particular interest since, unlike most of Earman's examples, it does not seem to involve, at least directly, any singularities, wild divergences, or any other tinkering with infinities of physically meaningful parameters in any way that often leave the true believer of determinism unsatisfied. Norton uses this example to support his vision of causality as a notion belonging more in folk science rather than being a fundamental principle underlying all natural processes and unifying all the domains of science at some deeper level.

While by and large sympathetic with the general thrust of the "anti-fundamentalist" programme of Norton, I intend to demonstrate that his mass on the dome example fails to provide support for such a view. I discuss the physical significance of an often overlooked and yet important Lipschitz condition, the violation of which underlies the existence of anomalous non-trivial solutions in this and similar cases. I argue that the singularity arising from the violation of the Lipschitz condition in the systems considered appears to be so fragile as to be easily destroyed by slightly relaxing certain (infinite) idealizations required by these models. In particular, I show that the idealization of an absolutely nondeformable, or infinitely rigid, dome appears to be an essential assumption for anomalous motion to begin; any slightest elastic deformations of the dome due to finite rigidity of the dome destroy the shape of the dome required for indeterminism to obtain. Furthermore, I demonstrate that this situation cannot be remedied by making the dome a little "pointier" at the apex, in the hope that the dome assumes just the right shape after it is "squished" down by the weight of the mass placed on top of the dome.

I also consider several further modifications of the original Norton's example – the rope-on-the-edge example and the rope-on-the-spherical-dome example – and show that indeterminism in these cases, too, critically depends on the nature of certain (infinite) idealizations pertaining to elastic properties of the bodies in these models. Most specifically, the idealization of an infinitely flexible rope appears to be an essential assumption for indeterminism to obtain; the rope of any finite degree of stiffness appears to be unable to follow the specific time-reversible shape of the underlying surface, thus blocking the time reversibility argument.

Admitting that the examples considered in this part in no way exhaust all possible relevant situations of interest, I argue that indeterminism of the Norton-type Lipschitz-indeterministic systems should perhaps be viewed as an artefact of certain (infinite) idealizations essential for the models, depriving the examples of much of their intended metaphysical import, as, for example, in Norton's antifundamentalist programme.

The rest of the Part I is organized as follows. Section 2.2.1 introduces Norton's original mass-on-the-dome example and the time reversal argument. In sections 2.2.2 – 2.2.5 I expose some of the loopholes of this example and introduce a cleaner version free of these unnecessary complications. Section 2.3 discusses the Lipschitz condition as it appears in the theory of ordinary differential equations (ODEs) and as it enters the mass-on-the-dome example. Section 2.4 is concerned with elastic phenomena that take place in the Norton-type indeterministic systems and that appear critical in the discussion of time-reversibility of their solutions. Here I also consider several further modifications of Norton's original example and show that certain idealizations required by these examples are so extreme as to be considered physically inadmissible. Section 2.5 applies (infinite) asymptotic reasoning to further elucidate the role and the domain of applicability of certain idealizations as they appear in philosophy of science. Finally, in Section 2.6 I draw several results from classical hydrodynamics to further illustrate how certain solutions associated with first-order differential equations with spatially non-Lipschitz velocity fields may lead to lack of important temporal properties of systems such as stability with respect to perturbations and Markovianity in time, and demonstrate how the behaviour of such systems may depend on the nature of the idealizations made.

2.2 The Mass on the Dome

2.2.1 Setting the Stage: The Original Formulation

A unit point mass slides frictionlessly on the surface under the action of gravity. The surface is shaped as a symmetric dome described by the equation:

$$y(r) = -\frac{2}{3g} r^{3/2}, \quad (1)$$

where r is the radial coordinate in the surface, i.e., the distance traveled by the mass from the highest point of the dome along the surface, $|y|$ specifies how far the dome surface lies below the apex as a function of r , and g is the acceleration due to gravity (Fig. 1):

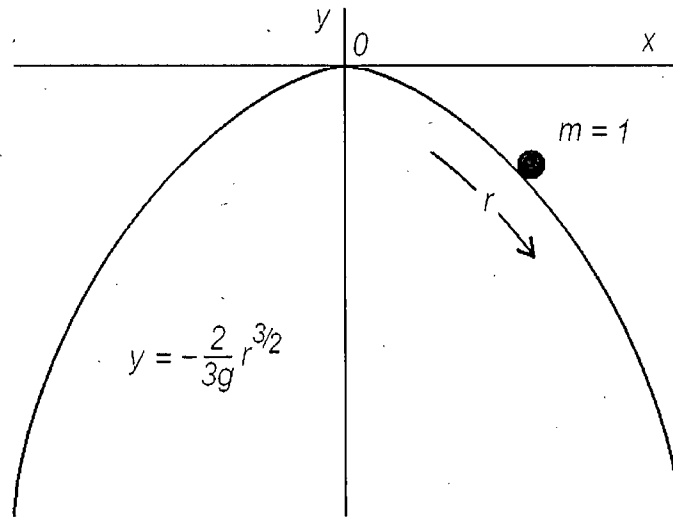


Fig. 1 Mass sliding on the dome.

At any point, the magnitude of the gravitational force tangential to the surface is

$$F_{\tau} = -\frac{d(gy)}{dr} = r^{1/2} \text{ and is directed outward. Newton's second law of motion, } F = ma,$$

applied to the mass on the surface gives

$$\frac{d^2r}{dt^2} = r^{1/2}. \quad (2)$$

If the mass is initially located at rest at the apex $r = 0$, then one obvious solution to (2) for all times t is a trivial one:

$$r(t) = 0. \quad (3)$$

The mass simply remains at rest for all times. However, there exists another large class of unexpected solutions. For any radial direction,

$$r(t) = \begin{cases} \frac{1}{144}(t-T)^4, & \text{for all } t \geq T \\ 0, & \text{for all } t \leq T \end{cases}, \quad (4)$$

where $T \geq 0$ is an arbitrarily chosen constant. By direct computation one can readily confirm that (4) satisfies Newton's second law (2).

Note that equation (4) describes a point mass sitting at rest at the apex of the dome, whereupon at an arbitrary time $T \geq 0$ it spontaneously moves off in some arbitrarily chosen radial direction.

The solutions (4) appear to be fully in accord with Newton's second and first laws, if one takes the first law in its instantaneous form as follows:

In the absence of a net external force, a body is unaccelerated.

Indeed, for all times $t < T$, there is no net force applied, since the body is at position $r = 0$, the force free apex; and the mass is unaccelerated.

For all times $t > T$, there is a non-zero net force applied, since the mass is at positions $r > 0$ *not* the apex, the only force free point on the dome; and the mass accelerates in accord with $F = ma$.

Finally, when $t = T$, the direct computation of the mass acceleration from the equation (4) gives us

$$a(t) = \begin{cases} \frac{1}{12}(t-T)^2, & \text{for all } t \geq T \\ 0, & \text{for all } t \leq T \end{cases}, \quad (5)$$

so that at $t = T$, the mass is still at the force-free apex $r = 0$ and the mass acceleration $a(0)$ is equal to zero. Again, no force, no acceleration, exactly as the first law requires.

What about the initiating cause that sets the mass in motion in the first place? Surely the instant $t = T$ is *not* the *first* instant at which the mass moves; it is the *last* instant at which the mass does *not* move. In fact, one can name no first instant at which the mass moves. So, if there is no first instant of motion, then there is no first instant at which to seek the initiating cause.

Yet another powerful argument can be given in support of acausality of the mass motion. This argument involves the time reversal trick. Since the Newtonian dynamical laws of gravitational systems are invariant under time reversal we can invert the sliding-down-the-dome scenario to produce another legitimate solution which insults the principle of causality. Instead of having the mass starting at the apex of the dome, we will imagine it starting at the rim and that we give it some initial velocity directed exactly at the apex. If we give it too much initial velocity, it will pass right over the apex to the other side of the dome. If we give it too small initial velocity, it will rise toward the apex, but before it reaches the apex it halts and slides back to the rim. Now, if we give the mass just the right amount of initial velocity, it will rise up and momentarily halt exactly at the apex.

The proper mathematical analysis of the latter situation reveals that the time required for the mass to reach the apex moving along the surface of this particular shape is *finite*. That this time is finite is essential for the time reversal trick to succeed. Infinite time would mean that the mass never actually arrives at the apex, and the time reversal scenario would display a mass that has been in motion at all past times, without any spontaneous launches. It should be emphasized that by no means this feature is common to all domes. For hemispherical or parabolic domes, for instance, the time taken for the mass to reach the apex to its momentary halt is unbounded. In the case of the dome of Fig. 1 the time reversal trick does work.

2.2.2 The Mass on the Dome: Cartesian Perspective

Defining the surface by $y(r) = -\frac{2}{3g}r^{3/2}$, in terms of the distance r traveled by the mass along the surface, conceals important details about the actual geometry of the dome and directionality of motion as viewed from the "external" Cartesian perspective. Special care should be taken not to overlook these details since they may prove crucial in the general case. Indeed, as the present section shows, the original formulation of the mass-on-the-dome example harbours several loopholes, some of which potentially fatal to the project. Fortunately, most of these difficulties are reparable and can be overcome by slightly modifying the original formulation of the problem. As the next sections show, the modified version inherits all the strangeness of being a source of spontaneous motion generation without our having to deal with the loopholes and unnecessary qualifications.

Consider an axially symmetric dome, and x is a Cartesian radial variable. In what follows, we consider only non-negative x 's, and then extend the results to incorporate the

negative values of the x -coordinate. For the curvilinear coordinate r measured along the surface slice cut vertically through the apex at the origin we have

$$\frac{dr}{dx} = \sqrt{1 + [y'(x)]^2},$$

so that

$$\frac{dy(r)}{dr} = -\frac{1}{g} r^{1/2} = \frac{dy}{dx} \cdot \frac{dx}{dr} = \frac{y'(x)}{\sqrt{1 + [y'(x)]^2}}, \quad y'(x) \leq 0.$$

Expressing the coordinate r we can write down

$$r = g^2 \frac{[y'(x)]^2}{1 + [y'(x)]^2}. \quad (6)$$

Putting $v = y'$ and differentiating both sides of (6) gives us

$$\frac{dv}{dx} = \frac{1}{2g^2} \frac{(1 + v^2)^{5/2}}{v}, \text{ so that } \int \frac{v dv}{(1 + v^2)^{5/2}} = \frac{1}{2g^2} x + \text{Const}.$$

Integrating the left-hand side of the last expression we get:

$$(1 + [y'(x)]^2)^{3/2} = \frac{1}{C - kx}, \quad (7)$$

for $k = \frac{3}{2g^2}$ and some constant C .

The first observation to make is that the dome surface appears to be defined not for all x 's, but only for x 's out of some (final) interval $[0, L)$. Indeed, the left-hand side of (7) is always greater or equal to 1, so it must be the case that $0 \leq C - kx \leq 1$ for all x out of some interval $[0, L)$. That provides the constraint on the possible values the constants C and L can take:

$$0 < kL \leq C \leq 1, \text{ where } L < 1/k.$$

Finally, integrating (7) and incorporating negative values of x , we can express $y(x)$ as a function of the Cartesian coordinate x :

$$y(x) = -\frac{1}{k}[1 - (C - k|x|)^{2/3}]^{3/2} + C_1, \quad (8)$$

where $0 < kL \leq C \leq 1$, $L < 1/k$, and the constant $C_1 = -\frac{1}{k}(1 - C^{2/3})^{3/2}$.

The tangential gravitational force acting on the mass as a function of x is

$$F_\tau = g \frac{y'(x)}{\sqrt{1 + [y'(x)]^2}}. \quad (9)$$

The normal force exerted upon the mass by the surface at the point x is

$$N = g \frac{1}{\sqrt{1 + [y'(x)]^2}}. \quad (10)$$

2.2.3 The Mass on the Dome vs. the Mass on the Pinnacle

Having expressed the shape of the dome in the linear Cartesian coordinates, it is easy to see that *not all* the domes described by formula (8) would fit well for generating spontaneous motion within Newtonian mechanics.

Indeed, having fixed some "rim", $L < 1/k$, and depending on the value of the constant C , two distinct cases are possible.

Case 1: $C \neq 1$. Substituting $x = 0$ into (7) we obtain $(1 + [y'(0)]^2)^{3/2} = \frac{1}{C} > 1$, so that the first derivative of the function tends to some non-zero constant, d , as x approaches zero. Geometrically this means that the tangent line to the dome surface at zero hits the y -axis at some non-zero angle – the mass arrives at the apex not exactly horizontally but at some non-zero angle. As we pass through zero into negative x 's, the tangent line to the surface experiences a sudden step-like jump:

$$\lim_{x \rightarrow 0+} y'(x) - \lim_{x \rightarrow 0-} y'(x) = 2d \text{ for some } d \neq 0.$$

So $y'(x)$ is simply not defined at $x = 0$. I shall refer to this case as the *mass-on-the-pinnacle* scenario (Fig. 2):

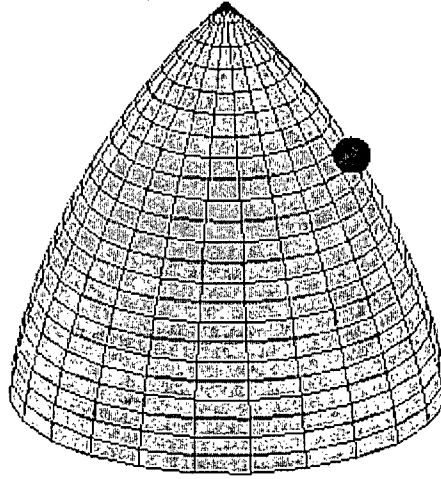


Fig. 2 The mass on the pinnacle.

Since $y'(x)$ enters the expressions (9) and (10) for the tangential gravitational and normal forces acting on the mass, these forces appear not to be defined at zero either. As Newton's second law of motion " $F = ma$ " cannot be written for the mass at zero, the mass-on-the-pinnacle scenario simply does not belong in Newton's mechanics jurisdiction, and should be excluded from the discussion by an appropriate stipulation. Yet, I shall return to this case again in section 2.3, where it appears in regard with the Lipschitz condition.

Case 2: $C = 1$. Since the constant $C_1 = -\frac{1}{k}(1 - C^{2/3})^{3/2}$, we have $C_1 = 0$, and the expression (8) for the shape of the dome takes a relatively simple form:

$$y(x) = -\frac{1}{k} [1 - (1 - k|x|)^{2/3}]^{3/2}, \quad (11)$$

for all x out of the interval $(-L, L)$, $L < 1/k$, for which the surface is defined.

Looking at (7), we see that the first derivative, $y'(x)$, turns to zero as x approaches zero. Geometrically this means that the mass arrives at the apex exactly horizontally. Even though the second and higher derivatives of $y(x)$ do diverge at zero, this fact by itself, at least directly, seems to generate no singularity or divergence in any physically meaningful parameter. I shall refer to this situation as the (proper) case of the mass on the dome (Fig. 3). All discussion that follows below will have this situation in mind.

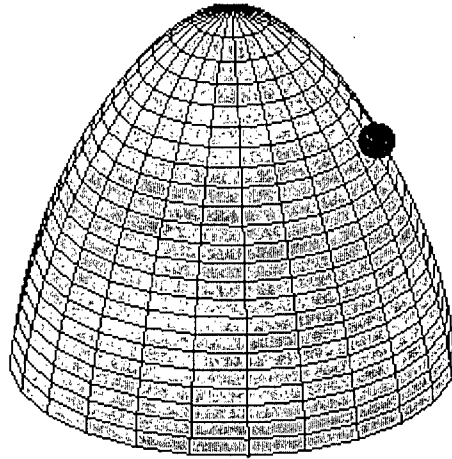


Fig. 3 The mass on the dome (proper).

At this point we also note that in the vicinity of the apex, i.e., at the limit $|x| \rightarrow 0$, the graph behaves as a fractional power function:

$$y(x) \approx -\sqrt{8k/27} |x|^{3/2} = -\frac{2}{3g} |x|^{3/2}, \quad (12)$$

which, of course, should have been expected, since, for small x 's, the x -coordinate almost coincides with the radial coordinate r measured along the practically horizontal surface. I

shall return to this point later in the next section when trying to modify the original formulation of the problem to avoid some of its loopholes.

On the other side of the interval, as x approaches the "rim", $|x| \rightarrow L$, the graph descends steeper and steeper until it hits the vertical wall at $|x| = L$. (The graph of this function appears in Fig. 1.)

2.2.4 The Mass on the Dome or the Mass in the Air?

Further observations about the behaviour of the mass on the so defined dome are in order. Consider again the three classes of trajectories produced if we give the mass at the rim some initial velocity directed at the apex along the surface: those where the mass halts before it reaches the apex and falls back to the rim; those where the mass halts exactly at the apex; and those where the mass passes the apex with some non-zero velocity and rushes over to the other side of the dome.

There is an easy and instructive way to see how things may go astray by considering the trajectories from the third class when the mass passes over the apex. As the mass proceeds through the apex with a *non-zero* velocity into negative x 's, it continues its motion along an artillery shell like ballistic trajectory. However, in the vicinity of the apex, the dome surface descends *faster* than any such parabolic trajectory so that the mass necessarily *detaches* itself from the surface once it enters the negative x 's, *refusing* to follow the prescribed track.⁸ (The results of numerical simulation for different velocities appear below in Fig. 4.)

⁸ The mass detachment for the third class of trajectories was also noticed by David Malament (manuscript).

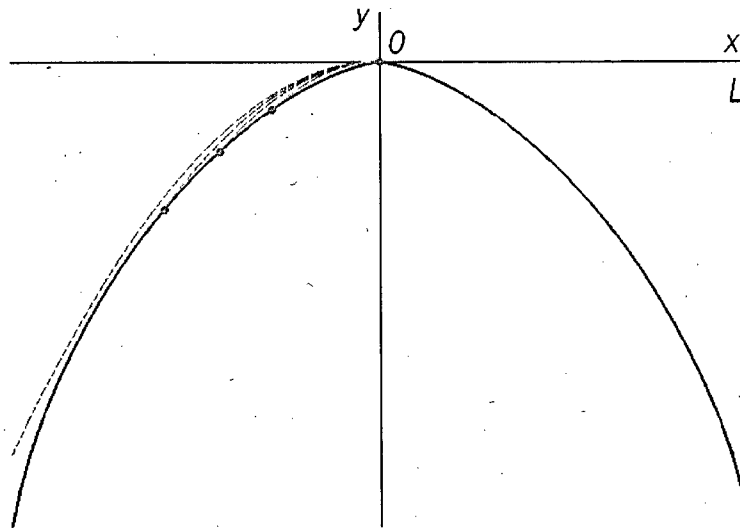


Fig. 4 Passing over the dome apex.

It is worth noting, however, that mass detachment for the *third* class of trajectories in no way affects Norton's original argument and is mentioned here for illustrative purposes only. More important is similar detachment that occurs in the *second* class of trajectories, for which the mass is aimed to halt exactly at the apex. To see this, recall that at the "rim", $x = L$, the tangential to the surface plane is exactly vertical. It means that the mass, initially positioned at the rim and given any initial velocity directed at the apex along the surface, will go up (and then fall back) precisely vertically, *detaching* itself from the surface and thus, again, *refusing* to follow its curvature. A careful analysis reveals that this is true not only for the rim $x = L$ taken as an initial position of the mass, but also for some vicinity of the rim. Indeed, for the so defined domes, there always exists a (finite) interval (L_1, L) , $0 < L_1 < L$, such that, for all initial positions of the mass (the "rims") taken within this interval, the ballistic trajectory of the mass descends *slower* than the dome surface immediately under it, thus causing the mass to detach from

the surface. (The results of numerical simulation for different initial positions are shown below in Fig. 5.)

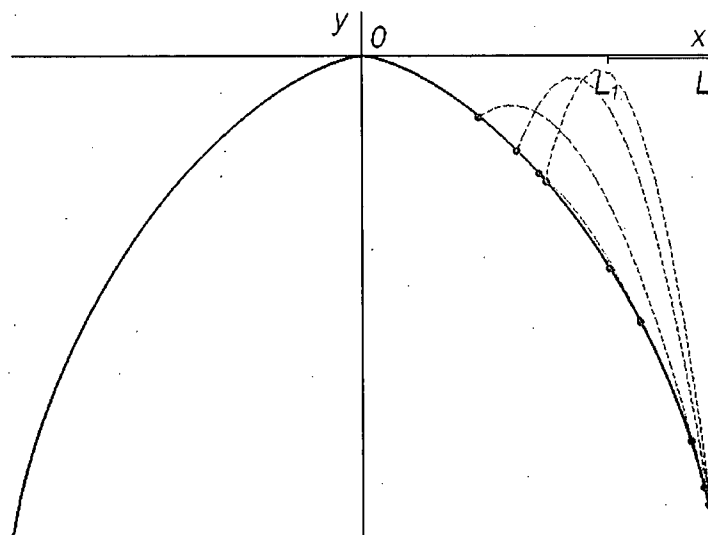


Fig. 5 Mass detachment in the vicinity of the rim.

Clearly, such initial positions would be a poor choice for being the "rim" since, as mentioned in section 2.2.1, letting the mass fly along its ballistic parabolic path would irreparably ruin the time reversal argument.

Any attempt to bypass this difficulty by pushing the mass back on the prescribed track (e.g., switching to a *bead-on-the-wire* example) would necessarily involve additional forces (*viz.*, the elasticity forces of the wire along which the bead would slide) without which the mass will simply refuse to follow the track. Adding new forces (external or internal), as we shall soon see, brings new and undesirable complications into Norton's original problem.⁹

⁹ This phenomenon of mass detachment is closely connected with a mechanical system's being *ideal holonomic* since an ideal holonomic constraint can be taken as the limiting case of a system with a large potential energy, or, equivalently, the limiting case of an *infinite* force field in a neighbourhood of the

These complications become especially important (and more subtle) when we move away from the rim to the vicinity of the apex. Unlike the previous situation, for any initial position of the mass taken within the interval $(0, L_1)$, the ballistic trajectory of the mass now descends *faster* than the dome surface, and *no* detachment of the mass from the surface occurs. (The results of numerical simulation are shown in Fig. 6.)

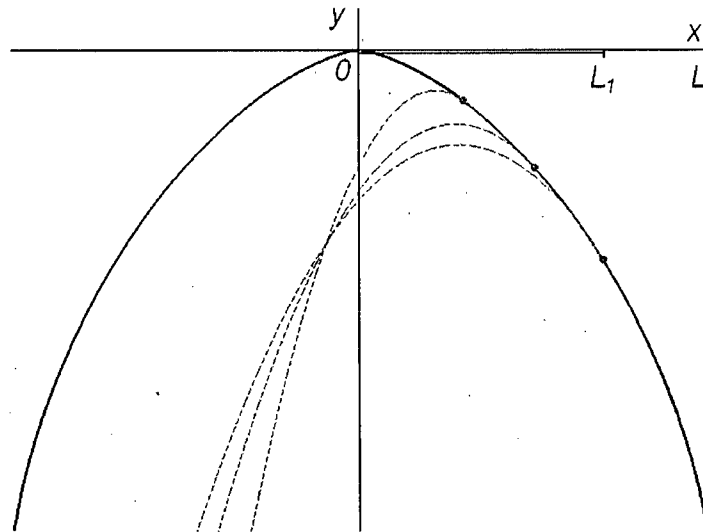


Fig. 6 No mass detachment in the vicinity of the apex.

Though we need not resort to elastic wires going through the mass to keep it from detaching itself from the surface of the dome, this is the elasticity of the dome that acquires special importance here.

At this point we can discern a general pattern that begins to emerge. We don't want to let the mass move along its free-flight parabolic trajectory; parabolic trajectories give infinite past times for the time reversal scenarios, thus stripping the whole argument of its force. In the vicinity of the rim this can be done by invoking the elasticity forces.

curve, directed toward the curve to ensure the moving point remains exactly on the curve (see Arnold 1978). I intend to expand on this point elsewhere.

of the wire *through* which the bead-mass now slides (or applying some other external force). In the vicinity of the apex these are the elasticity forces of the dome that would "straighten up" the mass path sufficiently to yield the required curvature; were it not for the elasticity forces of the dome, the mass would choose to follow its free-flight ballistic path. In any case, there is a (very strong) force field in the neighbourhood of, and directed toward the surface, that ensures that the mass moves along the required path. Yet, as the following sections show, *no* finite, however large, elasticity coefficient of the dome can allow for the time reversibility of the mass motion; only *absolutely* rigid dome can make the time reversal trick possible; thus the singularity in a physically meaningful parameter of the situation.

2.2.5 The Mass on the Dome, Modified

Many of the difficulties mentioned above can be overcome and will disappear if we make the following minor change in the original formulation of the problem. Namely, instead of defining the surface by $y(r) = -\frac{2}{3g}r^{3/2}$, in curvilinear terms of the distance r traveled by the mass along the surface, we define it, just as (12) suggests, in the usual linear Cartesian coordinates by

$$y(x) = -\frac{2}{3g}|x|^{3/2}. \quad (13)$$

This way, first, the situation no longer harbours the mass-on-the-pinnacle case in which the slope of the surface experiences a sudden step-like jump at the apex. The mass moving along the surface will now always arrive at zero exactly horizontally and no qualifications to the contrary are necessary.

Second, the surface is now defined for *all* x 's, not just for all x out of some interval $(-L, L)$, $L < 1/k$; no more "preferred" "rims" or vertical walls.

Third, with any arbitrary (x -positive) point on the dome taken as the motion starting point (the "rim"), no detachment of the mass from the surface when it starts at its "rim" ever occurs; no need to resort to elastic wires to keep the mass on the prescribed track.

On the other hand, since at the infinitesimal vicinity of the apex (the spontaneous motion generation region) the curvilinear coordinates and linear coordinates coincide, the simplicity of the expression (2) for the second law of Newton remains in place.

This does not, however, prevent the mass from detaching from the surface for the third class of trajectories once the mass passes the apex with non-zero velocity as in Fig. 4. Fortunately, these trajectories play no role in Norton's time reversal argument, so we will simply let the mass disappear from our attention once it vanishes behind the apex into the other side.

As the following sections show, the so modified mass on the dome example inherits all the strangeness of being a source of spontaneous motion generation with no need to deal with the above discussed loopholes and unnecessary qualifications.

2.3 The Lipschitz Condition

We recall that the function $x(t)$ satisfying the initial condition (t_0, x_0) is a solution of the differential equation determined by a vector field v

$$\frac{dx}{dt} = v(x) \quad (13)$$

if the following identity holds for all t in the interval I on which $x(t)$ is defined:

$$\frac{dx}{dt} = v(x(t)), \text{ and } x(t_0) = x_0. \quad (14)$$

Every differential equation (13) defines a direction field of this equation in the plane: the line attached at the point (t, x) has slope $v(x)$. If x_0 is a singular point of the vector field, i.e., $v(x_0) = 0$, then $x(t) = x_0$ is a solution of the equation (13) satisfying the initial condition $x(t_0) = x_0$. Such a solution is called an *equilibrium position* or *stationary solution*.

Let $v(x) = \operatorname{sgn}(x) |x|^{3/4}$. For such a field, (13) has more than one solution, e.g., the solutions $x_1(t) = 0$ and $x_2(t) = |t/4|^4$ satisfy the same initial condition $(0, 0)$. In fact, (13) has a whole 1-parameter family of solutions obtained by gluing together the corresponding halves of the two solutions, $x_1(t) = 0$ and $x_2(t) = [(t - T)/4]^4$, at some arbitrary time $T > 0$. (This situation is typical in that if (13) has more than one solution, then it has a "continuum" (i.e., a closed connected set) of solutions.) In the general theory of ordinary differential equations it is hardly a surprising fact; if the direction field v is continuous but nondifferentiable (only Hölder continuous), the solution with initial condition in the equilibrium position may fail to be unique. Indeed, for any $v(x) = \operatorname{sgn}(x) |x|^\alpha$, where $0 < \alpha < 1$, there always exists a family of branching solutions for (13) that satisfy the same initial condition $(0, 0)$:

$$x(t) = \pm[(1 - \alpha)(t - T)_+]^{1/(1-\alpha)}, \quad (15)$$

where $f_+ = \max(f, 0)$, for an arbitrary T (Fig. 7):

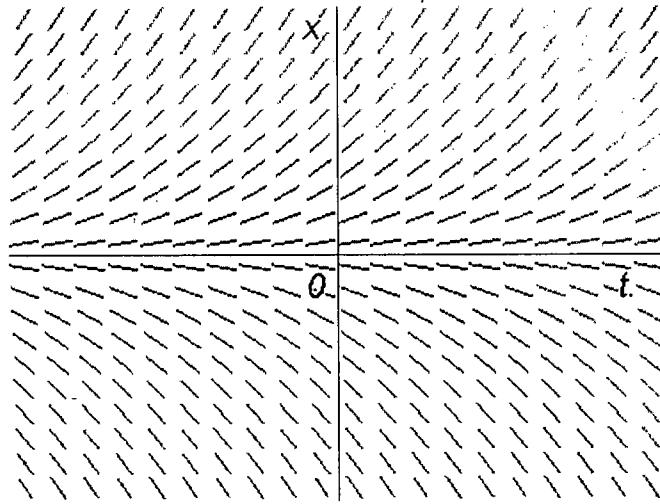


Fig. 7 The direction field for $\alpha = 3/4$.

Geometrically, the reason for non-uniqueness in these cases is that the velocity decreases too slowly when approaching the equilibrium position. As a result, the solution manages to reach the singular point in a finite time. It turns out that the smoothness of v guarantees the uniqueness in these cases. This observation needs more elaboration.

Let us assume that $x(t)$ is a solution of the equation $\frac{dx}{dt} = v(x)$ with a smooth right-hand side v . We will suppose that $x(t_0) = x_0$ is an equilibrium position and $x(t_1) = x_1$ is not (Fig. 8):

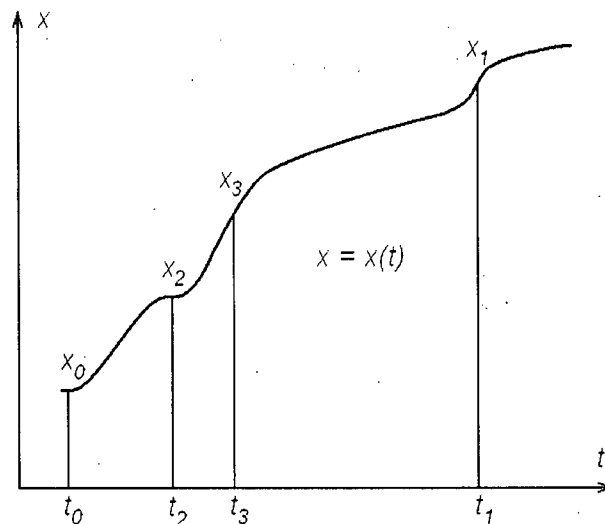


Fig. 8

On the interval between t_0 and t_1 , consider the instant t_2 closest to t_1 such that $v(x(t_2)) = 0$. By Barrow's formula for any point t_3 between t_2 and t_1 we have

$$t_3 - t_1 = \int_{x_1}^{x_3} \frac{d\xi}{v(\xi)}, \quad x_3 = x(t_3).$$

If the function v is smooth, then the integral tends to infinity as x_3 tends to x_2 . Indeed, the slope of the chord of the graph of a smooth function on an interval is bounded, so that $|v(\xi)| \leq k |\xi - x_2|$, where the constant k is independent of the point ξ of the interval $[x_1, x_2]$ (Fig. 9):

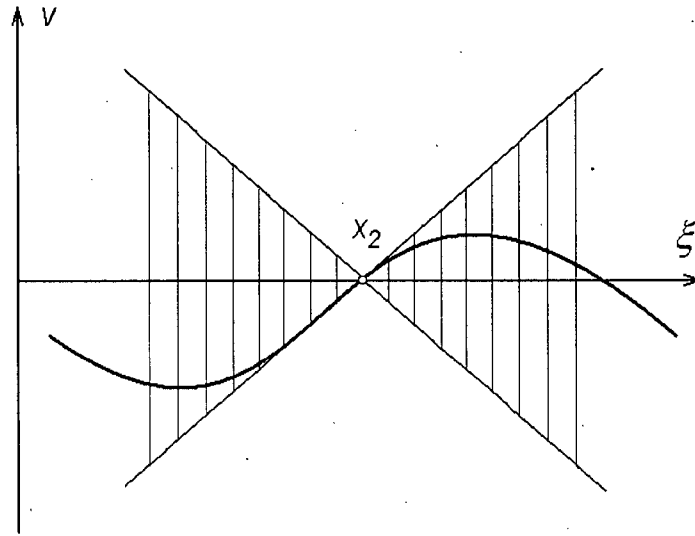


Fig. 9

Thus

$$|t_3 - t_1| \geq \left| \int_{x_1}^{x_3} \frac{d\xi}{k(\xi - x_2)} \right|.$$

The latter integral is easily calculated; it tends to infinity as x_3 tends to x_2 . It is easy to verify this without even calculating the integral: it must be equal to the time of transit between the two points in the linear field (i.e., for parabolic trajectories), and this time (logarithmically) tends to infinity when one of the points tends to the equilibrium position.

Thus the number $|t_2 - t_1|$ is larger than any pre-assigned number. So the solution with initial condition in an equilibrium position cannot assume values that are not equilibrium positions. Therefore if $x(t_0)$ is an equilibrium position, we have $v(x(t)) \equiv 0$

for all t . Consequently $\frac{dx(t)}{dt} \equiv 0$, i.e., $x(t)$ is a constant. The uniqueness is now proved.¹⁰

Note that the main point of the proof was the comparison of a motion in a smooth field with a more rapid motion in a suitable linear field (i.e., *parabolic* trajectories). For the latter motion the time to enter an equilibrium position is infinite, and consequently it is *a fortiori* infinite for the slower motion in the original field. Indeed, it can be shown that a sufficient condition for uniqueness of the solution with initial value x_0 is that the

integral $\int_{x_0}^x \frac{d\xi}{v(\xi)}$ diverge at x_0 .

The condition that $|v(\xi)| \leq k |\xi - x_2|$, where the constant k is independent of the point ξ of the interval $[x_1, x_2]$ (i.e., the condition that the slope of the chord of the graph be bounded), is called a *Lipschitz condition* and the constant k a *Lipschitz constant*.¹¹ It can be shown that a sufficient condition for uniqueness is that the right-hand side function v satisfy a Lipschitz condition $|v(x) - v(y)| \leq k |x - y|$ for all x and y .

It is by no accident that we chose the function $|v| = |x|^{3/4}$ to exemplify the Lipschitz discontinuity. Writing down the energy conservation relation for a unit mass sliding on the surface of the axially symmetric dome defined by the equation $y(x) = -|x|^{3/2}$ gives us

¹⁰ The proof is due to Arnold (1992).

¹¹ More generally, if $|f(\xi) - f(x)| \leq k |\xi - x|^\beta$ for given x and all $|\xi - x| < \delta$, where k, β are independent of ξ , and $\beta > 0$, and α is the upper bound of all the β such that a finite k exists, $f(\xi)$ is said to satisfy a Hölder condition (or, in some textbooks, a Lipschitz condition of order α) at $\xi = x$. If a Lipschitz condition of order α is satisfied at x , it can be shown that $f'(x) = 0$. If at every point of the interval it is satisfied for some $\alpha > 1$, then $f'(x) = 0$ throughout the interval. Hence $f(x)$ is constant; consequently only the case $0 < \alpha \leq 1$ is of primary interest.

$$\frac{\dot{x}^2}{2} = g |y(x)|,$$

so that

$$\left| \frac{dx}{dt} \right| = \sqrt{2g} |y(x)|^{1/2} = \sqrt{2g} |x|^{3/4},$$

which, to the factor of $\sqrt{2g}$, coincides with our $|v|$. The right-hand side of the expression is non-Lipschitz: its derivative, $\frac{3\sqrt{2g}}{4} |x|^{-1/4}$, is unbounded in the neighbourhood of the origin. As a result, the uniqueness theorem does not apply. Gluing together in a smooth manner the corresponding halves of the two solutions, $x_1(t) = 0$ and $x_2(t) = [(t - T)/4]^4$, at some arbitrary (positive) time T reproduces Norton's anomalous solutions. That's the reason why the existence of anomalous non-trivial solutions in Norton's case can be wholly attributed to (spatial) Lipschitz-discontinuity of the system's velocity field (i.e., Lipschitz-discontinuity of the square root of the potential well wherein the mass moves). Let us call such solutions (*spatial*) *Lipschitz indeterministic*.

A further observation clarifying the role of the Lipschitz condition is in place. A Lipschitz condition is weaker than that for the function v to be C^1 (continuously differentiable). Indeed, the uniqueness theorem holds in the case when the first derivative of v exists but is discontinuous. What that means with respect to our situation is that the mass-on-the-pinnacle case cannot be *amended* by *defining* the otherwise undefined values of the derivative $y'(x)$. Try, for instance, to define it as the right-hand limit of the derivative at zero (i.e., regard the mass at apex positioned as if it is still on the right-hand side of the surface only):

$$y'(0) := \lim_{x \rightarrow 0^+} y'(x) < 0.$$

Now that the derivative exists but is discontinuous, the uniqueness theorem comes into play, and only one solution survives. Which one and why? Substituting this value of $y'(0)$ into (9) we see that, at the apex, there is always a non-zero tangential gravitational force F_t pushing the mass downward. It is to say that, after having reached the apex and having momentarily halted, the mass will necessarily turn back and fall down to where it came from. The situation becomes no more paradoxical than throwing a stone vertically in the air, seeing it halt after some finite time, and then catching it back again. The trivial solution $x_1(t) = 0$ is impossible on so defined pinnacle surface exactly for the same reasons that make it impossible to have the stone hang in the midair forever once it has reached its maximal altitude. (A similar argument applies if one tries to define $y'(0)$ as the left-hand limit of the derivative at zero.) Causality reigns.

2.4 Elastic Deformations and Physically Inadmissible Idealizations

2.4.1 Models and Idealizations

Scientific theories employing mathematical models "approximate" or "idealize" in one way or another. Whereas much attention in philosophy of science has been drawn to the role of idealizations in the development of scientific theories (e.g., McMullin 1985, Cushing 1990, Moulines 1996, see also Redhead 1980, Laymon 1985, 1995, and Hartmann 2005a, 2005b), it is admissibility of idealizations in theorizing that will be of main interest in this section. I will argue that certain idealizations required by Norton's dome example are so extreme as to be considered physically inadmissible. In particular,

the idealization of absolutely nondeformable, or infinitely rigid, dome appears to be an essential assumption for indeterminism to obtain; any slightest diversion from this idealization that allows any (however small!) elastic deformations of the dome completely destroys the shape of the surface at the apex needed for spontaneous motion generation to occur. In addition, there seems to be no way of remedying this situation by starting out with another elastic dome surface that is a little "pointier" at the apex, in the hope that the dome assumes just the right shape after it is "squished down" by the weight of the mass placed on top of the dome. As a result, indeterminism of the dome example should rather be viewed as an artefact of certain (infinite) idealizations, depriving the example of much of its intended metaphysical import, as, for example, in Norton's antifundamentalist programme.

2.4.2 Elastic Deformations and Idealization of a Concentrated Force

In a real physical body that is not deformed, the arrangement of the molecules corresponds to a state of thermal equilibrium; all parts of the body are in mechanical equilibrium. This means that, if some portion of the body is considered, the resultant of the forces on that portion is zero. When a change in the relative positions of molecules – a deformation – occurs, the body ceases to be in its original state of equilibrium. As a result, there arise forces which tend to return the body to equilibrium. These internal forces which occur when a body is deformed are called *internal stresses*. If no deformation occurs, there are no internal stresses.

The internal stresses in a real physical body are due to molecular forces, i.e., the forces of (electro-magnetic) interaction between the molecules. The molecular forces tend to have a very short range of action: their effect extends only to the neighbourhood

of the molecule exerting them, over a distance of the same order as that between the molecules. On the other hand, the standard classical theory of elasticity, as a macroscopic theory, considers only distances which are large compared to the distances between the molecules. This being so, in many problems the atomistic structure of the elastic medium can often be disregarded and the body replaced with by a continuous mathematical model whose geometrical points are identified with material points of the medium.

The internal forces determine the elastic properties of bodies, which mathematically characterize certain functional relationships between forces and deformations of elastic medium. As a result, the response of elastic body to the action of forces (internal or external) is in no way arbitrary, but is subject to certain relationships and constraints that may prove critical in many problems.

Consider the following problem from the classical theory of elasticity. Suppose we are given a vertically symmetric dome of the (already familiar) form:

$$y(x) = -|x|^{3/2}.$$

The dome is made of infinitely divisible continuous elastic medium, and the standard assumptions of the classical theory of elasticity are assumed to be in place.¹² We want to determine the deformation of the dome under the action of a finite concentrated force, applied vertically to just one point of the surface – the apex of the dome:

¹² See, e.g., Sokolnikoff (1956) and Thomas (1961). In particular, if a medium particle, initially at the point x_i , is displaced to the position x'_i at time t according to the relations $x'_i = \phi_i(x_i, t)$, these relations are assumed to have a unique inverse at any time t , ϕ_i must be continuous and differentiable functions of the initial coordinates x_i and the time t , and the functional determinants $\|\partial x'_i / \partial x_i\|$ and $\|\partial x'_i / \partial x_i\|$ are different from zero everywhere.

To solve this problem, we introduce polar coordinates, with an angle φ measured from the direction of the applied force (Fig. 10):

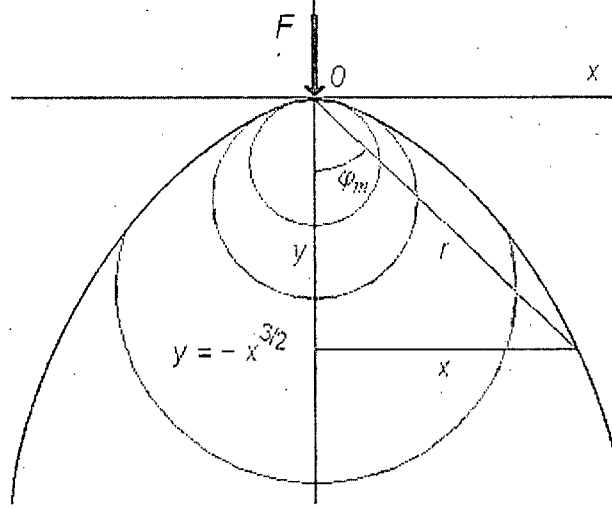


Fig. 10 Deformation of the dome under the action of a concentrated force.

For any given radial distance from the apex r , the angle φ takes values from $-\varphi_m$ to φ_m , where

$$|\tan \varphi_m| = |x/y| = |x|^{-1/2}.$$

The components of the stress tensor, σ_{rr} , $\sigma_{\varphi\varphi}$, and $\sigma_{r\varphi}$, can be expressed (in polar coordinates r, φ) as derivatives of the stress function, χ (Landau and Lifshitz 1986, p. 21):

$$\sigma_{rr} = \frac{1}{r} \frac{\partial \chi}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \chi}{\partial \varphi^2}, \quad \sigma_{\varphi\varphi} = \frac{\partial^2 \chi}{\partial r^2}, \quad \sigma_{r\varphi} = -\frac{\partial}{\partial r} \left(\frac{1}{r} \frac{\partial \chi}{\partial \varphi} \right).$$

Since at every point of the dome boundary except the apex where the force is applied we have

$$\sigma_{\varphi\varphi} = \sigma_{r\varphi} = 0,$$

the stress function χ should satisfy the following conditions:

$$\frac{\partial \chi}{\partial r} = f_1(\varphi), \quad \frac{1}{r} \frac{\partial \chi}{\partial \varphi} = f_2(\varphi),$$

for $\varphi = -\varphi_m, \varphi_m$, and some functions f_1 and f_2 not depending on r .

Substituting $\chi(r, \varphi) = rf(\varphi)$, the biharmonic equation of equilibrium for elastic solid bodies (Landau and Lifshitz 1986, pp. 18, 48)

$$\Delta \Delta \chi = \left[\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \right) + \frac{\partial^2}{\partial \varphi^2} \right]^2 \chi = 0$$

gives solutions for $f(\varphi)$ of the following forms:

$$\sin \varphi, \cos \varphi, \varphi \sin \varphi, \text{ and } \varphi \cos \varphi.$$

The first two of these correspond to stresses equal to zero identically, and it is only the third one that gives the correct value for the force applied at the apex. Indeed, projecting the internal stresses on directions parallel and perpendicular to the force \mathbf{F} , and integrating over the part of a small circle lying inside the dome and centered at the apex, in the limit of zero radius we obtain

$$\int \sigma_{rr} r \cos \varphi d\varphi = -F,$$

$$\int \sigma_{rr} r \sin \varphi d\varphi = 0,$$

as required to balance the external force applied at the apex. From here we can get the following solution:

$$\chi(r, \varphi) = -\left(\frac{F}{\pi}\right) r \varphi \sin \varphi,$$

$$\sigma_{rr}(r, \varphi) = -\left(\frac{2F}{\pi r}\right) \cos \varphi,$$

$$\sigma_{\varphi\varphi} = \sigma_{r\varphi} = 0,$$

where F is the force per unit thickness of the dome.

These formulas determine the required stress distribution in the dome. Note that the stress distribution is purely radial: only a radial compression force acts on any area perpendicular to the radius. The lines of equal stress are the circles $r = d \cos \varphi$, which pass through the apex and whose centres lie on the line of the action of the force F .

The components of the strain tensor are

$$u_{rr} = \sigma_{rr}/E, \quad u_{\varphi\varphi} = -\sigma\sigma_{rr}/E, \quad \text{and} \quad u_{r\varphi} = 0,$$

where σ is Poisson's ratio and E is Young's modulus characterizing the elastic medium of the dome.

Now, expressing the components of the strain tensor in terms of the derivatives of the components of the displacement vector (in spherical polar coordinates):

$$u_{rr} = \frac{\partial u_r}{\partial r}, \quad u_{\varphi\varphi} = \frac{1}{r} \frac{\partial u_\varphi}{\partial \varphi} + \frac{u_r}{r}, \quad 2u_{r\varphi} = \frac{\partial u_\varphi}{\partial r} - \frac{u_\varphi}{r} + \frac{1}{r} \frac{\partial u_r}{\partial \varphi},$$

we can get the final expression for the displacement vector that solves our problem:

$$u_r = -\frac{2F}{\pi E} \log(r/a) \cos \varphi - \frac{(1-\sigma)F}{\pi E} \varphi \sin \varphi,$$

$$u_\varphi = \frac{2\sigma F}{\pi E} \sin \varphi + \frac{2F}{\pi E} \log(r/a) \sin \varphi + \frac{(1-\sigma)F}{\pi E} (\sin \varphi - \varphi \cos \varphi).$$

Here the constants of integration have been chosen so as to give zero displacement (translation and rotation) of the dome as a whole: an arbitrary point at a distance a from the apex on the line of action of the force is assumed to remain fixed (Fig. 11):

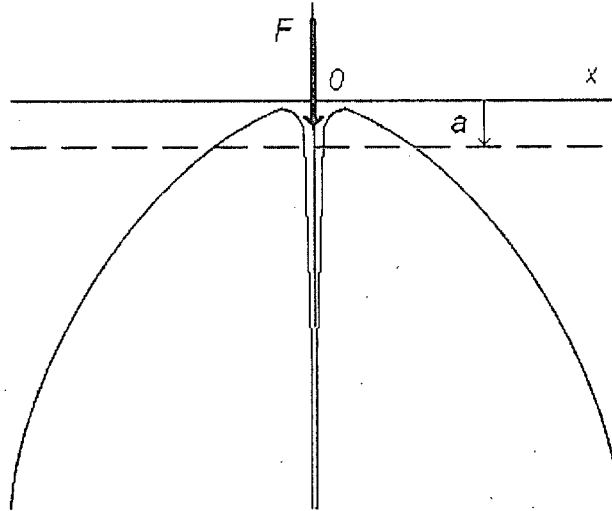


Fig. 11 Logarithmic infinite well at the apex.

Looking at the formal solution just given we notice that it presents an *infinitely deep logarithmic well* going all the way down from the apex. Assuming that the mass always remains in contact with the surface on which it exerts the force (i.e., the mass travels down with the surface as the latter is "squished" down by the mass), it follows that there can be no trivial solution with the mass *staying* on (or sliding along) the surface – the only shape of the dome that counterbalances a concentrated finite force exerted by a point mass is the infinite logarithmic well.

That we get an infinitely deep well (as opposed to some finite "pimple" on the surface) is not, perhaps, a totally surprising result: it is characteristic for this particular model that the force is applied to just one point – the area of measure zero, thus producing *infinite pressure* on the surface at the apex. Disallowing any fractures and punctures of the surface which is taken to remain continuous at all times (yet infinitely stretchable), the elastic surface will always give way under the infinitely sharp needle as the latter pushes its way down. That is all to say that the problem of a point mass staying

on top of an elastic surface is *not* a well-posed problem in the first place – it cannot even be set up properly within the classical theory of elasticity.

2.4.3 Modification 1: The Dome with a Pinnacle on Top

The infinite logarithmic well also helps explain why the original Norton's mass-on-the-dome formulation cannot be *remedied* by making the elastic dome surface a little "pointier" at the apex, in the hope that the surface will assume just the right shape after the mass is placed on top of the dome and "squishes" it down by its weight. Indeed, consider, for example, the following modification of the dome with an arbitrarily high and sharp conical pinnacle on top (with the non-zero angle 2α as close to zero, and the height of the pinnacle, h , as high as desired) (Fig. 12):

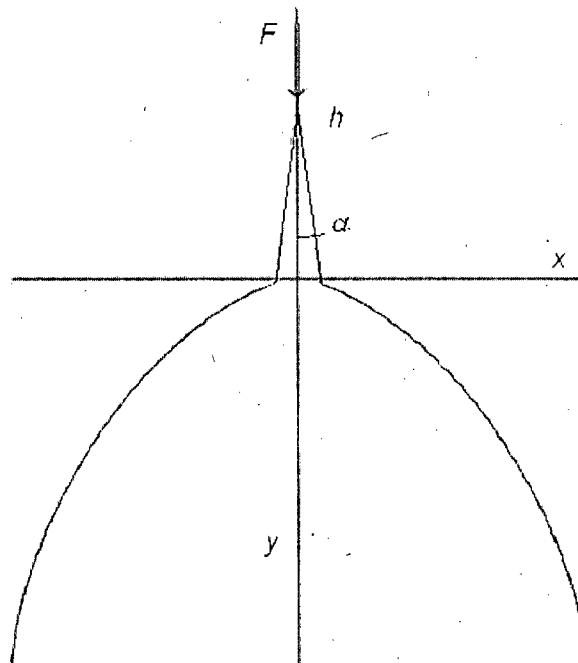


Fig. 12 The dome with a pinnacle on top.

The stress distribution in such a pinnacle-shaped part of the dome due to a concentrated force applied vertically to its apex is, obviously, given by the same formulae as above with only difference being in their normalization constants. Namely, the stress tensor components in this case are (cf. Landau and Lifshitz 1986, pp. 48-49):

$$\sigma_{rr}(r, \varphi) = - \left(\frac{F}{\alpha + \frac{1}{2} \sin 2\alpha} \right) \frac{1}{r} \cos \varphi, \quad \sigma_{\varphi\varphi} = \sigma_{r\varphi} = 0.$$

Such stress distribution, however, also gives rise to an infinitely deep logarithmic well going all the way down from the top of the dome. Since this is true for any arbitrarily high and sharp pinnacle, there is no way the new dome can assume the desired shape at the apex after we place a point mass on its top.

2.4.4 Modification 2: The Rope-Sliding-the-Edge Example

One may think that all the pathologies of the Norton's dome originate in the singularity of the dome surface at the apex. Indeed, in the original Norton's setting it is the singular non-smooth geometry of the surface in the immediate vicinity of the apex that is essential for indeterminism to obtain; the rest of the surface plays no role in this phenomenon and can be safely replaced by some other surface, or even merely removed. Yet, as I will show in this and the following sections, the singularity of the surface at the apex is in no way essential for indeterminism to occur and one can get anomalous motion generation for everywhere smooth surfaces.

Recall that the primary resource responsible for non-uniqueness in Norton's case is (spatial) Lipschitz-discontinuity of the direction field of the differential equation that governs the system's dynamics. For a simple Newtonian gravitational system such as

Norton's dome it is equivalent to Lipschitz-discontinuity of the square root of the potential wherein the mass moves. Correspondingly, we can expect similar indeterminism in systems whose dynamics is governed by the same differential equation, as long as the square root of their *effective potential* is non-Lipschitz.

Consider the following example.¹³ A flexible yet unstretchable rope of negligible thickness lies motionlessly on a frictionless flat horizontal surface with one of its ends touching the edge of the surface; beyond this point the surface descends abruptly with the shape given by the equation expressed in usual Cartesian coordinates (Fig. 13):

$$y(x) = -|x|^{1/2}.$$

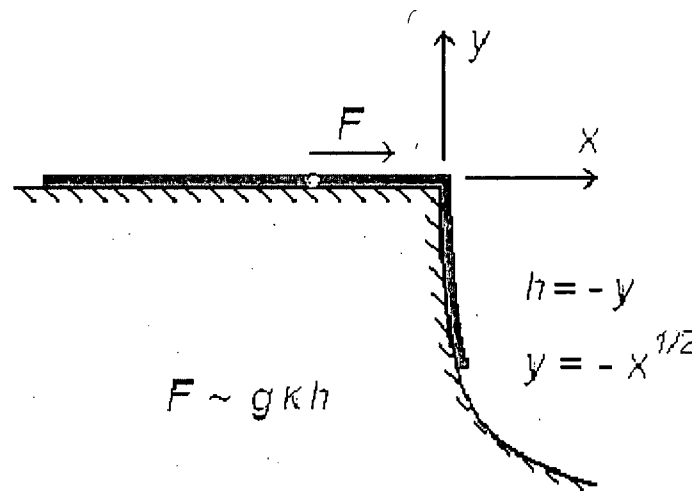


Fig. 13 The rope sliding down the edge.

When the rope slides down the edge by some distance x (in the x -direction), the force exerted on any given point of the rope is

$$F \sim g\kappa h = g\kappa |x|^{3/2},$$

¹³ This example was suggested by William Unruh (personal communication).

where κ is the rope mass per length unit, and h measures the distance from the x -axis to the sliding end of the rope. Equivalently, we may talk about the centre of mass of the rope moving in the potential

$$U(x) = -g\kappa |x|^{3/2}.$$

The rope moving in this effective potential, clearly, presents a similar dynamical situation as that of Norton's point mass on the dome, giving rise to similar indeterministic behaviour of the rope: for an arbitrarily long time the rope lies motionlessly on the horizontal surface, when, suddenly, without any external intervention or change in the environment, it starts sliding down the edge.

The fact that the rope-sliding-the-edge example has no point masses makes it of special interest. No point masses, no concentrated forces, no infinite pressures on the surface, and no infinite wells that would prevent the problem from being well posed in the first place.

Also note that it is the rope's centre of mass the behaviour of which is governed by the non-Lipschitz differential equation, and this centre of mass is now an abstract *mathematical* point, not necessarily corresponding to any *material* point of the rope. Indeed, once the rope slides off its original position, its centre of mass detaches from the rope and moves to the depth of the medium that constitutes the slide. That the system's centre of mass no longer necessarily tracks the path of the system itself is an important feature that will allow us to further modify the example as to smoothen all singularities in the surface while maintaining indeterministic non-Lipschitz dynamics.

Yet, before we do so, it is instructive to look more closely at the elastic phenomena taking place in the immediate vicinity of the edge where the rope is bent

before it goes down the wall. More specifically, I will argue that the idealization of an infinitely flexible rope in this situation is essential for indeterminism to occur; taking into account any slightest elasticity effects (however small!) destroys any non-trivial solution resulting in the rope's remaining motionless at all times with no spontaneous motion generation.

2.4.5 Bending of Rods

Consider an elastic rope (or a rod) of finite, yet negligible, thickness. Assuming finite stiffness of the rope, let us look carefully at its bending in the immediate vicinity of the edge (Fig. 14):

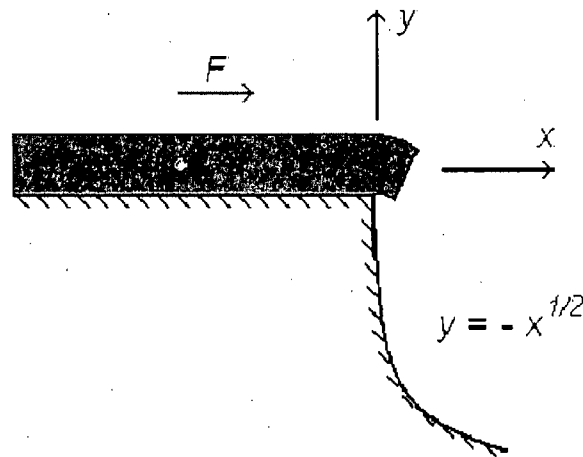


Fig. 14 The bending of the rope in the immediate vicinity of the edge.

When a rope (or a rod) of non-zero thickness is bent, it is stretched at some points and compressed at others. Lines of the convex side of the bent rod are extended, and lines on the concave side are compressed. This being so, there is a neutral surface in the

rod that undergoes neither extension nor compression. This neutral surface separates the region of compression from the region of extension.

Let us look at a bending deformation in a small portion of the length of the rod in the immediate vicinity of the edge where not only the strain tensor but also the magnitude of the displacements of points in the rod are assumed small. Taking a coordinate system with the origin on the neutral surface in the portion considered, and the x -axis parallel to the axis of the undeformed rod, we can suppose that the bending occurs in the xy -plane. Not giving a general solution for all possible configurations of the system, we give the expressions for the components of the displacement for a rod of rectangular cross-section as it appears, e.g., in Landau and Lifshitz (1986, pp. 65-66):

$$u_x = xy/R, \quad u_z = -\sigma yz/R, \quad \text{and}$$

$$u_y = -\frac{1}{2R}\{x^2 + \sigma(y^2 - z^2)\},$$

where R is the radius of curvature of the neutral surface at the edge.

Using these expressions, it can further be shown that the sides $z = \pm \frac{1}{2}z_0$ of the initially rectangular cross-section become $z = \pm \frac{1}{2}z_0 + u_z = \pm \frac{1}{2}z_0(1 - \sigma y/R)$, i.e., no longer parallel but still straight. The sides $y = \pm \frac{1}{2}y_0$, however, are bent into the *parabolic* curves

$$y = \pm \frac{1}{2}y_0 + u_y = \pm \frac{1}{2}y_0 - \frac{1}{2R}[x_0^2 + \sigma(\frac{1}{4}y_0^2 - z^2)].$$

That the bottom edge of the rope of any finite degree of stiffness is bent into *parabolic* curves makes it impossible for the rope to follow the underlying surface;

rope's refusal to follow the prescribed track once again blocks the time reversibility argument – the rope stays on the horizontal surface forever.

There is another way to see that a finitely flexible rod cannot be deformed as to make it stay in a close contact with the surface at every point when the surface experiences a sudden change in directionality such as in the above example: the displacements of the rod in the just given expressions diverge and cease to be well defined if the radius of curvature of the neutral surface R reaches zero. This also helps understand why an (initially straight and undeformed) rod with any finite degrees of stiffness, when placed on top of Norton's original dome, *detaches* from the dome in a manner closely resembling the detachment of the mass projectile when the latter passes over the top with a non-zero velocity (Fig. 15):¹⁴

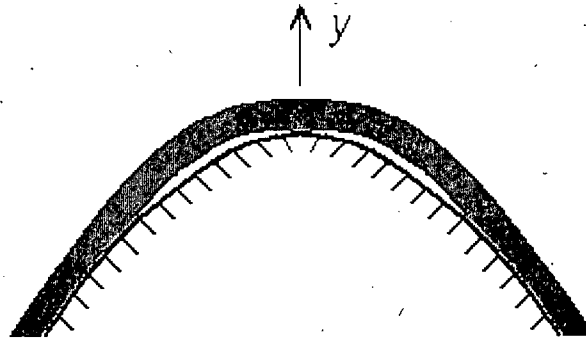


Fig. 15 The bending of the rope in the immediate vicinity of the apex.

Indeed, the radius of curvature R of Norton's original dome surface (or, more precisely, that of any two-dimensional cross-section of the surface cut vertically through the apex) is given by the following familiar expression of differential geometry:

¹⁴ Similar difficulties, it may be argued, may be expected when trying to force the mass to track the surface by switching to a bead-on-the-wire example; deformations of the wire at the origin would, on a microscopic level, presumably, diverge the bead unto a parabolic, not time-reversible, trajectory.

$$R = \frac{(1 + y'^2)^{3/2}}{y''},$$

where the two-dimensional curve is given explicitly by the equation $y = f(x) = -|x|^{3/2}$

(Fig. 16):

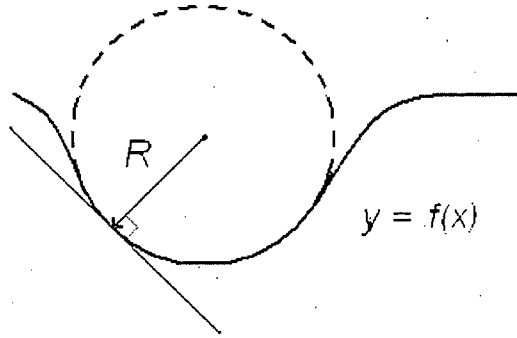


Fig. 16 The radius of the curvature of a plain curve.

This radius R (which, in the limit of infinitely thin rope lying on the dome, coincides with the radius of curvature of the rope's neutral surface), clearly, tends to zero when we approach the top of the dome.¹⁵

Yet, the next section will show that the situation can be improved still further as to eliminate all singularities in the geometry of the surface while maintaining indeterminism.

2.4.6 Modification 3: The Rope on the Spherical Top Dome

Consider the following modification of the rope-sliding-the-edge example. A flexible (of some finite degree of stiffness) yet unstretchable rope of negligible thickness lies motionlessly on a frictionless symmetrical dome. In some vicinity of the top the dome is

¹⁵ The observation that the Gaussian curvature of Norton's dome at the apex is infinite first appears in David Malament (manuscript). He seems, however, to think that this is merely a geometrical fact about the surface which is no more physically troublesome than other idealizations employed in Newtonian theory.

spherical; beyond this spherical area the dome continues in a smooth manner as the familiar power function (Fig. 17):

$$y(x) = -|x|^{3/2}.$$

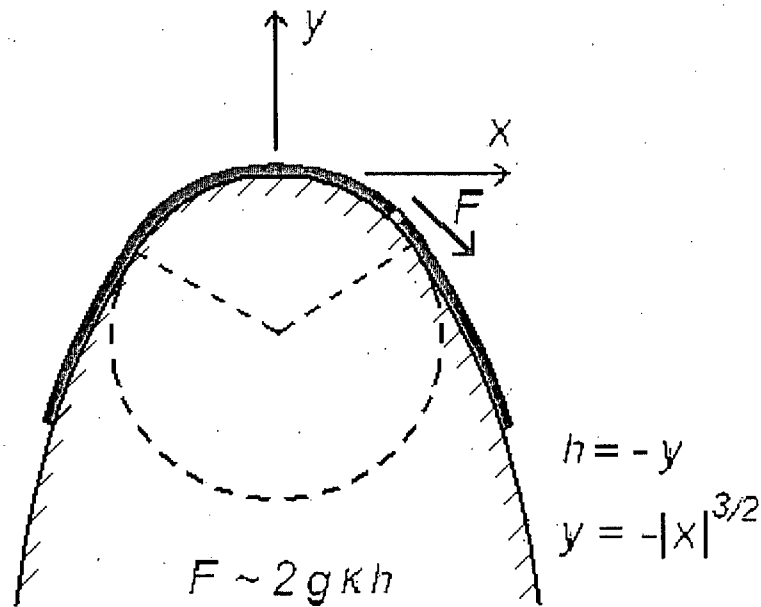


Fig. 17 The rope sliding down the spherical top dome.

The rope initially lies motionlessly on the dome with the two halves of the rope hanging symmetrically from both sides of the dome. The total length of the rope exceeds that of the spherical arc of the dome so that the rope's ends protrude to the non-spherical parts of the dome.

When the rope slides down the dome by some distance x (in the x -direction), the force exerted on any given point of the rope is

$$F \sim 2g\kappa h = 2g\kappa |x|^{3/2},$$

where κ , again, is the rope mass per length unit, and h measures the distance from the x -axis to the sliding end of the rope. Equivalently, we may talk about the centre of mass of the rope moving in the potential

$$U(x) = -2g\kappa |x|^{3/2}.$$

The rope moving in this effective potential (which, up to the constant 2, coincides with that of the previous problem), also harbours spontaneous motion generation: for an arbitrarily long time the rope lies motionlessly on the dome, when, suddenly, without any external intervention or change in the environment, it starts sliding down the dome.

Notice that the surface of the dome is now perfectly smooth on top – it is spherical – with no singularities in the geometry of its shape, and yet, the velocity field of the differential equation for rope's centre of mass (i.e., the square root of the effective potential in which the rope moves) is Lipschitz-discontinuous at $x = 0$ resulting in the system's evolving in a non-unique manner.

Notice also that no difficulties with detachment arise when the (initially unbent) rope is bent into a spherical arc when it slides over the spherical part of the dome (any circular curve behaves locally as a parabola). In addition, once bent into a spherical arc, the rope can easily slide along the spherical top of the dome with no further bending thanks to rotational (in the vertical plane) symmetry of the sphere. This does not, however, apply to the non-spherical parts of the dome where the surface is no longer rotationally symmetric, and where the rope, therefore, necessarily undergoes additional continuous bending as it moves forward.

Not providing a rigorous analysis for this problem for all possible configurations of the system, I will argue that, assuming finite stiffness of the rope, similar (as in the

previous example) detachment phenomena may well be expected: looked at the microscopic level, the bottom edge of the rope's free end, while locally bent into parabolic curves, detaches from the dome surface, unable to follow the underlying surface of the time-reversible shape. The condition of non-Lipschitz direction field essential for indeterminism to obtain appears to be so fragile as to be easily destroyed by taking into account any slightest elastic effects within the system. Indeterminism, once again, appears to be an artefact of infinite idealizations.

Before concluding this section, however, I want to point out that, though this pattern seems to be true for all the examples considered above, by no means these cases exhaust all possible Lipschitz-indeterministic systems in which masses are not concentrated at a point and which, correspondingly, cannot be taken as settled. For instance, William Unruh¹⁶ suggested another possible modification of the previous examples in which the rope of finite stiffness is replaced by a drop of water of negligible thickness; the drop is lying on a surface of one of the above described shapes waiting to slide down. This drop-of-water-on-the-dome example, though lacking the difficulties with staying in contact with the surface at all times, presumably, presents a much harder case to analyse. In particular, it seems to involve a number of additional hydrostatic and hydrodynamic phenomena (such as surface tension and fluid transport), about particular models (and the assumption thereof) of which one should be very careful. As a possibility, it may be argued that it is an essential feature of this model (and, perhaps, of many of spatially extended models) that all elastic deformations must propagate throughout the elastic medium with infinite speed so that every part of the drop instantly "feels the tug" of any other part of the drop when the latter suddenly comes to motion. If

¹⁶ In personal communication.

so, the singular nature of certain assumptions essential for such models would, too, be revealed. Yet, for the time being, I leave this example unsettled and open for future research.

2.5 Asymptotic Reasoning in Philosophy of Science

2.5.1 Introduction

In his recent book Robert Batterman (2002) discusses what he calls asymptotic reasoning in physics, i.e., the qualitative analysis of behaviour of physical theories in the neighbourhood of singular limits, and its relevance to philosophical issues of explanation, reduction and emergence. Batterman argues that many physically and philosophically important theories and models involve a new and powerful category of explanation based on asymptotic reasoning that has been totally overlooked by philosophers of science.

Whereas much of Batterman's efforts have been directed on issues of inter-theoretical reduction and the development of the new type of explanation based on the idea of emergence, (infinite) asymptotic reasoning can prove helpful in elucidating the role and domain of applicability of various idealizations used the Norton-type Lipschitz-indeterministic models considered above, and, therefore, better understanding the extent of metaphysical import that these models can offer for today's philosophical debates on the nature of scientific determinism.

2.5.2 The Fallacy of Infinite Reasoning

One of the key ideas involved in infinite asymptotic reasoning is a commonplace fact of mathematics that *finite* and *infinite* compositions may differ in their essential properties.

For example, a finite intersection of open sets is always an open set, but an infinite intersection of open sets can yield a closed set. If some property is preserved at any finite stage of a finite sequence of operations, there is no guarantee that this property will be preserved in the transition to the infinite stage. One may even argue that many paradoxes in philosophy can be traced down to the so called fallacy of infinite reasoning – the fallacy of incorrectly projecting properties from finite to infinite compositions.¹⁷

Another example illustrating the fallacy of infinite reasoning is the following famous "proof" that $2 = \sqrt{2}$ (Fig. 18):

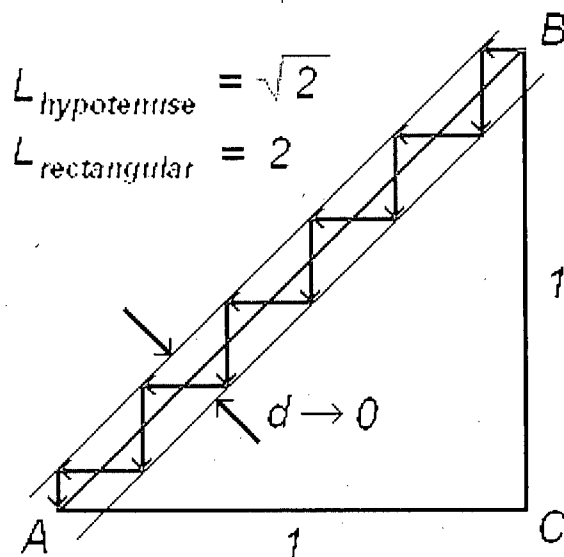


Fig. 18 The "proof" that $2 = \sqrt{2}$.

Here we have a rectangular triangle ABC with the unit-length sides and the hypotenuse equal to $\sqrt{2}$. Now, instead of moving from the point B to point A along the hypotenuse, we decide to move in a crooked rectangular manner: the hypotenuse surrounded with a tube of some (small) width, d , we start moving straight to the left

¹⁷ See, e.g., Earman and Norton (1996) for applying this reasoning to argue against known supertask paradoxes.

(parallel to the side AC) until we hit the tube's border. Then we turn 90 degrees counterclockwise and continue moving straight down (parallel to the side BC) until we hit the tube's border again. Then we turn 90 degrees clockwise and continue moving straight to the left until the next encounter with the tube's border. We then continue moving in this manner until we reach the point A (the width of the tube is assumed to be such that we can do it). It is easy to see that the total distance traveled along the crooked rectangular path is equal to 2, and yet, in the zero limit of the width of the tube, the crooked path diverges from the hypotenuse by no more than any arbitrarily small preassigned number.

2.5.3 Stability and Parameter Sensitivity

This phenomenon can be seen as a particular instance of a more general pattern encountered routinely in such mathematical disciplines as the system stability and control theory, the theory of parameter sensitivity in dynamical systems, catastrophe theory, and robotics.

Unlike the more traditional approaches in which to determine the properties of a system has been to exhibit a complete set of exact solutions of the equations describing this system, and then to study the properties of these solutions, in catastrophe theory it is realized that in many instances it is only information of a *qualitative* nature, or only limited quantitative information, which is the ultimate goal of the study of some systems of equations. In such cases a full spectrum of solutions to an equation, obtained by much hard work (if at all), may be a hindrance rather than a help in understanding the *qualitative* properties of the equation or system of equations (Gilmore 1981).

As a part of mathematics, catastrophe theory is a theory about singularities. Many interesting phenomena in nature (or, rather, their mathematical models) involve some discontinuities – breaking of a wave, the division of a cell or the collapse of a bridge. When applied to scientific theories, it deals with the properties of discontinuities directly, without reference to any specific underlying mechanism. This makes it especially appropriate for the study of systems whose inner workings are not known, or too complicated, and for situations in which the only reliable observations are of discontinuities.

In robotics, to maintain system stability has been the prime concern when designing any practical machine. There always exists a certain discrepancy between an actual (real-operating) and the nominal (theoretical) trajectories of any system. This discrepancy is partly due to various inherently approximational schemes in system identification, and partly due to possible further parameter variations stimulated by environmental changes. Thus, special attention should be paid to the evaluation of possible system parameter variations, and their effects on system's functional performance or "output" (see, e.g. Eslami 1994).

To illustrate the concepts of stability and parameter sensitivity let us consider a familiar classical system – a pendulum. A simple gravity pendulum – a weight on the end of a rigid rod, which, when given an initial push, swings back and forth under the influence of gravity over its central (lowest) point. As is known, the oscillations (not necessarily small) of the "ideal" pendulum are described by the following system of differential equations:

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = -\omega^2 \sin x_1, \quad \omega = \sqrt{l/g},$$

where x_1 is the angle of deviation from the vertical, x_2 is the angular velocity, l is the length of the pendulum, and g is the acceleration due to gravity.

The corresponding vector field in the phase plane with coordinates x_1, x_2 is just

$$v_1 = x_2, \quad v_2 = -\omega^2 \sin x_1,$$

with singular points $x_1 = m\pi, x_2 = 0$ (Fig. 19):

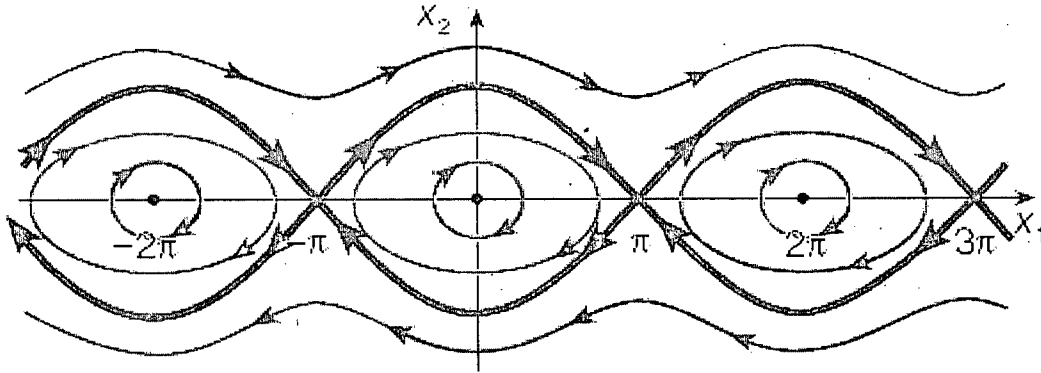


Fig. 19 Phase-space of a simple gravity pendulum.

If we restrict the motion of the pendulum to a relatively small amplitude, i.e., $|x_1| \ll 1$, the solution is a well-known harmonic oscillatory function:

$$x_1(t) = x_0 \cos(t/\omega), \quad x_2(t) = -(1/\omega) x_0 \sin(t/\omega),$$

where x_0 is the largest angle attained by the pendulum.

Period of the small oscillations is

$$T_0 = 2\pi/\omega.$$

For amplitudes beyond the small angle approximation, the exact period cannot be evaluated in terms of elementary functions and can only be written in the form of the elliptic function of the first kind:

$$T = \frac{4}{\omega} E\left(\frac{\sin x_0}{2}, \frac{\pi}{2}\right),$$

where $E(\psi, \varphi)$ is Legendre's elliptic function of the first kind:

$$E(\psi, \varphi) = \int_0^\varphi \frac{1}{\sqrt{1 - \psi^2 \sin^2 \xi}} d\xi.$$

The value of the elliptic function can also be computed numerically by using the following series:

$$T = T_0 \left[1 + \left(\frac{1}{2}\right)^2 \sin^2\left(\frac{x_0}{2}\right) + \left(\frac{1.3}{2.4}\right)^2 \sin^4\left(\frac{x_0}{2}\right) + \left(\frac{1.3.5}{2.4.6}\right)^2 \sin^6\left(\frac{x_0}{2}\right) + \dots \right].$$

A number of assumptions are built in this model: the bob of the pendulum is a point-mass; the rod on which the bob is swinging is massless and absolutely rigid; motion occurs in a vertical plane; there is no air resistance and friction at the nail, the material of which the bob is made is irrelevant to the study of the question, acceleration due to gravity does not depend on the position of the bob; there is no gravitational influence of the nearby objects at the mass, etc. Taken seriously, many of these idealizations are plainly unphysical (or physically inadmissible) in that they can never be achieved in practice for principle reasons, but, of course, no one is tempted to think that this "unphysicality" is indispensable to the relevant theory or that the theory would be absolutely unworkable without them.

A typical way how such conceptual "frauds" are dealt away in the teaching of science can be illustrated by the following excerpt:

One of the fundamental concepts of mechanics is that of a [*material point*].

By this we mean a body whose dimensions may be neglected in describing its motion. The possibility of so doing depends, of course, on the conditions of

the problem concerned. For example, the planets may be regarded as [material points] in considering their motion about the Sun, but not in considering their motion about their axes. (Landau and Lifschitz 1976, p. 1)

Implicit in such stipulations are our intuitions about the *stability* (or *robustness*) of the behaviour of the system with respect to disturbances or changes made to various parameters of the system. That is this feature of a system to change its behaviour insignificantly when the various parameters of the system are changed insignificantly that legitimizes some physical features to be idealized, or "neglected", as in the example above. This point needs more elaboration.

We recall that an equilibrium point x_0 of a system of differential equations,

$$\frac{dx}{dt} = v(x(t)), \quad (16)$$

is *locally Lyapunov stable* at $t = t_0$ if all solutions of this equation which start near x_0 (i.e., with their initial conditions in a neighbourhood of x_0) remain near x_0 for all time, i.e., if for any $\varepsilon > 0$ there exists a $\delta(\varepsilon, t_0) > 0$ such that

$$\text{if } \|x(t_0) - x_0\| < \delta \text{ then } \|x(t) - x_0\| < \varepsilon, \text{ for all } t \geq t_0,$$

for some appropriate choice of the norm $\|\dots\|$.

The equilibrium point x_0 is said to be *locally asymptotically stable* if x_0 is locally stable (in the sense of Lyapunov) and, furthermore, all solutions starting near x_0 tend towards x_0 as $t \rightarrow \infty$. Thus, the pendulum has a locally stable equilibrium point (but not asymptotically stable) when the pendulum is hanging straight down and an unstable

equilibrium point when it is pointing straight up. (If the pendulum is damped, the stable equilibrium point is locally asymptotically stable.)

Let us frame this situation differently in the following terms. Instead of disturbing the initial conditions of the system, we can talk of changing (smoothly) the total system energy taken as a parameter, and observe the *qualitative* changes in the behaviour of the system. As we start with small system energies (i.e., the initial conditions are around the point $(0, 0)$ in the phase-space diagram), the trajectories of the pendulum bob are closed curves; the pendulum performs back-and-forth oscillations. Increasing, in a smooth manner, the total energy of the system, we will eventually reach a point when, suddenly, the pendulum bob ceases to track a closed curve in the phase-space; in fact, it ceases to move at all once it takes an (unstable) vertical position with the zero velocity, and remains in this unmovable state ever since. This position is unstable – any, however light, disturbance of the pendulum returns it to the back-and-forth oscillatory motion.

If we increase the total energy even more, the pendulum starts a rotating motion, corresponding to non-closed curves in the phase-space diagram, resembling more and more straight lines as we push the energy still up. Thus, depending on the (non-zero) value of the chosen parameter (system's total energy), there exist three distinct behaviours of the system – that of a back-and-forth oscillations (closed phase trajectories), that of a halted (though unstable) position (a phase trajectory is just a single point), and that of a rotational motion about the axis of the pendulum (non-closed phase trajectories).

Depending on a particular problem in question, any other parameter of the system may be chosen to be disturbed or manipulated. An interesting (and important to us) case

is when we choose to manipulate the rod's elasticity coefficient k , and see whether the behaviour of the system will be, in some appropriate sense, robust under these perturbations.

Suppose, for example, that we are given a system of differential equations describing the behaviour of a gravitational pendulum, in which the elasticity of the rod is not assumed infinite but appears explicitly. Suppose further that, in this system of equations, the elasticity coefficient k of the rod is replaced by a new coefficient $k' = k + \delta k$; $x(t)$ is a solution of the original equation, and $x'(t)$ is a solution of the equation with the changed k' . For $x(t)$ to be a robust solution, we could require that for any appropriate disturbances δk there exists a $\delta(\varepsilon, t_0) > 0$ such that

$$\|x'(t) - x(t)\| < \delta, \text{ for all } t \geq t_0.$$

This is the latter sense of robustness that is essential for the possibility of the "unphysical" models' being used in simulating the behaviour of physical systems. Thus, by a *pendulum with infinitely rigid rod* we could now understand a *series* of (physically legitimate) *approximations* to the original problem, with finite but arbitrarily large and ever increasing elasticity coefficients, k 's, as long as

- (i) the series converges (in some appropriate sense) to some limiting solution (e.g., if the modulus of any two consecutive solutions can be made arbitrarily small by further increasing the corresponding elasticity coefficients), *and*
- (ii) this limiting solution has the same essential properties that any finitely approximate solution has.

2.5.4 The Lipschitz-Indeterministic Systems and Physically Inadmissible Idealizations

What does this have to do with the Norton-type Lipschitz-indeterministic examples considered above? Starting off with the original Norton's mass-on-the-dome example incorporating elastic phenomena, the behaviour of this system can be shown non-robust under ever increasing rigidity of the dome in the following sense: unless the stiffness of the dome is assumed *infinite*, the problem is not even a well-posed problem within the standard theory of elasticity; no domes of any finite (yet arbitrarily large!) stiffness can accommodate infinite pressures exerted by point masses placed on the surface.

The further modifications of the original Norton's example considered above can also be shown non-robust under ever increasing flexibility of the rope in the following, and more interesting, sense. These cases are all the more interesting that they closely resemble the structure of the above mentioned "proof" that $2 = \sqrt{2}$, harbouring the fallacy of infinite reasoning. Let flexibility of the rope is characterized by some coefficient of flexibility k . Take any finite, yet arbitrarily large flexibility coefficient k_1 of the rope. Place the rope in its initial position (either on the horizontal surface at the edge or on the spherical top dome) and wait whether it ever starts sliding. If it does, as we saw earlier, its bottom edge can only be bent (locally) into parabolic, non-time-reversible, curves. That is to say that, while macroscopically the rope appears to follow the shape of the dome, it actually (at a microscopic level) travels along a more complex path consisting of a number of shorter time-irreversible logs, thus blocking the time-reversal argument.

One can try to double (triple, etc.) the flexibility coefficient k_1 to get a new, larger, k_2 . It is obvious; however, that *qualitatively* the new situation is no different from the previous one with a less flexible rope. In particular, the new system, too, has only trivial solution (the rope is at rest at all times) and no spontaneous motion generation. Now, proceeding in a similar manner by making the rope more and more flexible, one gets an (infinite) sequence of systems with ropes of ever increasing (yet finite) coefficients of flexibility. Yet, as long as rope's coefficient of flexibility stays finite (and no matter how large), there is no spontaneous motion generation when the rope lies on the top.

The situation changes *qualitatively* if one takes an *infinitely* flexible rope; then (and only then) could the rope follow the particular shape of the dome as required by the condition of Lipschitz-discontinuity. Thus, the *limiting* behaviour of the family of approximations with finite rope's flexibility exhibits a certain property (of being non-unique), whereas any system in the approaching series (corresponding to a physically realistic, or admissible, situation) fails to exhibit this property. The only way to generate non-Lipschitz spontaneous motion of the rope is to allow the rope to be infinitely flexible; any diversion from actual infinity in the flexibility coefficient results in the rope staying on top forever.

There is an important lesson to be learned from these cases. As the Norton-type indeterministic examples discussed above show, there may exist models the real metaphysical power of which critically depends on the nature of certain idealizations made in those models, and the techniques of asymptotic reasoning may prove crucial in elucidating these issues.

2.6 Lipschitz-Indeterministic Solutions and the Markov Condition

2.6.1 Generalized Flows in Hydrodynamics

In this section I draw several results from classical hydrodynamics to further illustrate how certain solutions associated with first-order differential equations with spatially non-Lipschitz velocity fields may lead to lack of important temporal properties of systems such as stability with respect to perturbations and Markovianity in time, and show how the behaviour of such systems may depend on the nature of the idealizations made.

Consider the following transport equation for the scalar field $\theta(x, t)$ in $(x, t) \in \mathbf{R}^d \times [0, \infty)$:

$$\frac{\partial \theta}{\partial t} + (v(x, t) \cdot \nabla) \theta = 0, \quad \theta|_{t=0} = \theta_0. \quad (17)$$

In the classical theory of partial differential equations it is known that if the velocity field $v \in \mathbf{R}^d$ is bounded and continuous in (x, t) and Lipschitz continuous in x , then (17) can be solved uniquely by the method of characteristics. Denote by $\varphi_{s,t}(x)$ the solution of

$$\frac{d\varphi_{s,t}(x)}{dt} = v(\varphi_{s,t}(x), t), \quad (18)$$

starting at x at time s , i.e., with the initial condition $x(s) = x$. The solution of (17) is then given by the following expression:

$$\theta(x, t) = \theta_0(\varphi_{0,t}^{-1}(x)) = \theta_0(\varphi_{t,0}(x)). \quad (19)$$

The map $\varphi_{s,t} : \mathbf{R}^d \mapsto \mathbf{R}^d$ satisfying the following four properties:

- (a) $\varphi_{s,s}(x) = x$ for all s ;
- (b) $\varphi_{s,t}(x)$ is continuous in s, t, x ;
- (c) $\varphi_{\tau,t}(\varphi_{s,\tau}(x)) = \varphi_{s,t}(x)$ for all s, t, τ and all x ;
- (d) $\varphi_{s,t}(x) : \mathbf{R}^d \mapsto \mathbf{R}^d$ is a homeomorphism for all s, t

is called a *flow of homeomorphism* or, in short, a *flow*.¹⁸

These classical results are inapplicable when v , though bounded and continuous in (x, t) , fails to be Lipschitz continuous in x . In such cases no standard flow satisfying (a)–(d) can be associated with the ODE in (18) since the solution of this equation may fail to be unique. Since the solutions of (18) typically branch (i.e., (18) have more than one solution for the same initial condition), no forward-in-time map can be associated with such solutions. Similarly, no backward-in-time map can be associated with the solutions of (18) because they may coalesce on each other in finite time. This situation is unfortunate since transport in non-Lipschitz velocity fields may be physically motivated, e.g., for the problem of turbulence.¹⁹

Formally, the standard way to deal with this situation in general case is to randomize the set of maps which can be associated with the solutions in (18) by selection at the branching points thus defining a random field. Then a *generalized flow* associated with (18) can be defined as a random field with parameter (s, t, x) constructed by assigning a probability measure on the set of all maps associated with the solutions of

¹⁸ Exposition is due to Weinan E. and Vanden-Eijnden (2003).

¹⁹ The classical theory of Kolmogorov (1941) predicts that the solution of the Navier-Stokes equation in three dimensions is only Hölder with exponent $1/3$ in the limit of zero viscosity.

(18).²⁰ The generalized flows obtained in this way are typically non-degenerate random fields, i.e., the measure is not concentrated on a single point, due to branching.

As far as the modeling of the underlying physical processes is concerned, to pick the probability measure and single out *physically relevant* generalized flows the following *regularization procedure* is used. Instead of the original problems described by (17) and (18), the *regularized problems* with unique solutions are considered and the generalized flows are obtained as limits of the standard (stochastic) flows associated with these regularized problems.

Consider first the regularization by *smoothing of the velocity* around the points of Lipschitz discontinuity (the ε -limit process). Here the original equation (17) is understood as the limiting equation for the following motivating (and physically legitimate) problem:

$$\frac{\partial \theta}{\partial t} + (v^\varepsilon(x, t) \cdot \nabla) \theta = k \Delta \theta, \quad \theta|_{t=0} = \theta_0, \quad (20)$$

where k is the molecular diffusivity and v^ε is a mollified version of v on the scales $|x| \ll \varepsilon$ (e.g., if v solves Navier-Stokes equation, ε is the characteristic length scale associated with the kinematic viscosity). Unlike the original transport equation, (20) has a unique solution if either k or ε are positive. The generalized flow is then taken as the limit as $\varepsilon \rightarrow 0$ of the stochastic flow associated with (20), provided this limit can be defined in a suitable way.

Secondly, some *Brownian motion* (the k -limit process) can be added to the dynamics in (18) to obtain a unique (stochastic) flow associated with the solutions of

²⁰ For a more rigorous definition of generalized flows and their properties see Appendix A, Weinan E. and Vanden-Eijnden (2003).

$$dx = v^\varepsilon(x, t)dt + \sqrt{2kd}\beta(t), \quad (21)$$

where $\beta(\cdot)$ is a d -dimensional Wiener process (Stroock and Varadhan, 1969, 1979). The fact that the term $\sqrt{2kd}\beta(t)$ regularizes (21) can be understood more intuitively as thermodynamical fluctuations (always present in any real physical system unless we freeze it to the absolute zero), with probability 1, kicking instantaneously out any path that happens to go to the points x for which the solution of (18) is non-unique, thereby resolving the ambiguity associated with these positions. The generalized flow can now be defined; similarly, as the limit as $k \rightarrow 0$ of the stochastic flow associated with (20), provided this limit is defined in a suitable way.²¹

Mixed limits where both smoothing of the velocity field and Brownian motion are used can be considered as well.

The limiting generalized flows obtained in this manner, however, appear to depend sensitively on the regularization procedure, and they are non-Markov for generic regularizations. The latter fact raises an interesting issue regarding the connection of the Lipschitz indeterministic solutions with the Markov condition.

2.6.2 Non-Lipschitz Velocity Fields, Regularizations, and the Markov Condition

Consider first the following ODE we met in Section 4 (with $\alpha = 3/4$ corresponding to Norton's original formulation):

$$\frac{dx}{dt} = \text{sgn}(x)|x|^\alpha, \quad x \in \mathbf{R}, \quad \alpha \in (0, 1). \quad (22)$$

As mentioned before, the set of solutions of this equation is given by

²¹ Typically, the ε -limit is a weaker limit than the k -limit in the sense that the regularization by smoothing is more subtle due to the lack of stability of solutions to perturbations and issues with the choice of appropriate convergence. See W. E and Vanden-Eijnden (2003) for more details.

$$x(t) = \pm[(1-\alpha)(t-T)_+]^{1/(1-\alpha)}, \quad (23)$$

where $f_+ = \max(f, 0)$, for an arbitrary T .

By resolving the ambiguity of where to map the point $x = 0$, the following family of forward maps $\varphi_{s,t} : \mathbf{R}^d \mapsto \mathbf{R}^d$ can be constructed:

$$\varphi_{s,t}^\omega(x) = \begin{cases} \text{sgn}(x)(|x|^{1-\alpha} + (1-\alpha)(t-s))^{1/(1-\alpha)}, & \text{if } x \neq 0 \\ f^\omega(\tau)((1-\alpha)(t-\tau)_+)^{1/(1-\alpha)}, & \text{if } x = 0 \end{cases}, \quad (24)$$

where $\tau = \inf(t \geq s : f^\omega(s) \neq 0)$. Each of the maps $\varphi_{s,t}^\omega$ is a weak form of a flow, a *quasi-flow*, satisfying only the following three properties:

- (a') $\varphi_{s,s}^\omega(x) = x$ for all $s \in [0, T]$;
- (b') $\varphi_{s,t}^\omega(x)$ is continuous in s, t, x ;
- (c') $\varphi_{\tau,t}^\omega(\varphi_{s,\tau}^\omega(x)) = \varphi_{s,t}^\omega(x)$ for all x and for all $s, \tau, t \in [0, T]$ with $s \leq \tau \leq t$;

By superimposing these quasi-flows and assigning a suitable probability measure a generalized flow can be defined.

In this particular case, the generalized flows, obtained as limits of the standard flows by regularization either via the k -limit process or via the ε -limit process, can be shown to be Markov in time. However, as in the next example, Markovianity is not a generic property of generalized flows.

Consider a further generalization of the previous example:

$$\frac{dx}{dt} = \text{sgn}(x)|x|^\alpha g(t), \quad x \in \mathbf{R}, \quad \alpha \in (0, 1), \quad (25)$$

where g is a bounded function. Some solutions of (25) branch at the origin $x = 0$ on the time intervals where $g(t) > 0$, and other collapse at $x = 0$ where $g(t) < 0$ (see Fig. 20):

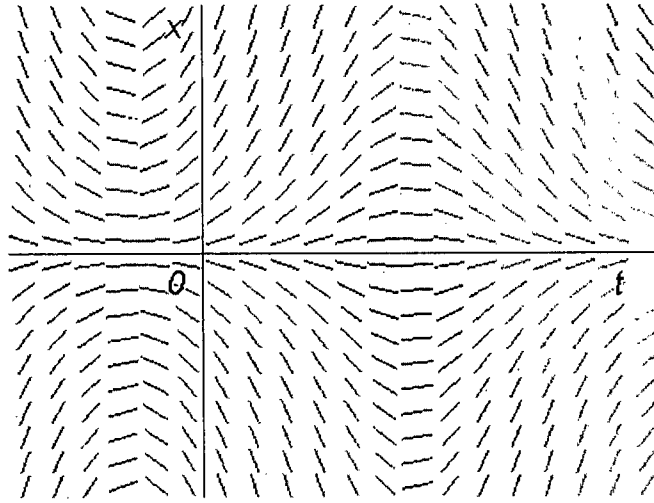


Fig. 20 The set of all solutions of (25) for $\alpha = 3/4$ and $g(t) = \cos(t)$.

Quasi-flows and the corresponding generalized flow have been properly defined, it can be shown that the generalized flows associated with (25) are not necessary Markovian in time. Though the generalized flows obtained by regularization by the k -limit process are Markov, the generalized flows obtained via the ε -limit process are not, so that this feature appears to depend sensitively on the regularization procedure used.

Finally, consider a further relaxation of the condition on the velocity field. In the following ODE the velocity field is continuous in (x, t) but non-Lipschitz at $(x, t) = (0, 0)$ and not even Hölder continuous $(x, t) = (0, 0)$:

$$\frac{dx}{dt} = v(x, t), \quad (x, v) \in \mathbf{R} \times \mathbf{R}, \quad (26)$$

where $v(x, t)$ is given by

$$v(x, t) = \begin{cases} \frac{2x}{t}, & \text{if } |x| > t^2, \\ 2 \operatorname{sgn}(x) \cdot t, & \text{if } |x| \leq t^2. \end{cases} \quad (27)$$

The set of solutions of (27) can be parameterized as

$$x(t) = \text{sgn}(a)t^2 + a, \quad x(t) = bt^2 \quad (28)$$

with $a > 0$, $b \in [-1, 1]$ (see Fig. 21):

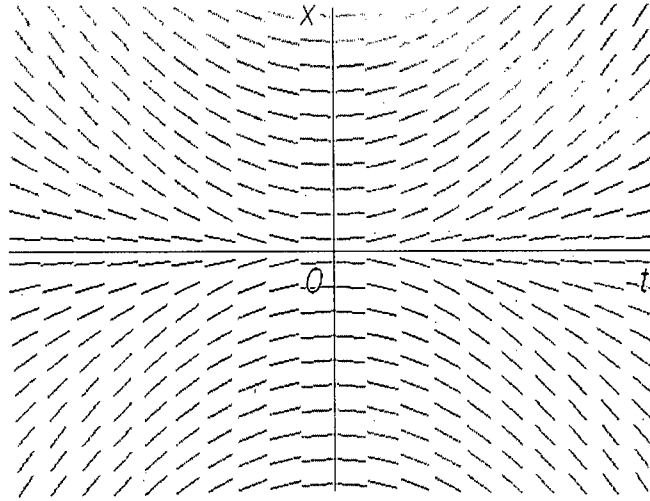


Fig. 21 The set of all solutions of (26).

In this case, the generalized flows associated with (26) can be shown to be non-Markov in time for both the k -limit process flows and most of the ε -limit process flows. The only regularization procedures that do produce a Markov generalized flows in the ε -limit process are those for which, in the limit $\varepsilon \rightarrow 0$, all the paths that collapse on a single node $(0,0)$ exit on a single trajectory.

To summarize the main points of this section, the purpose of drawing these results from fluid dynamics is to illustrate how properties specific to generalized flows associated with first-order differential equations with spatially non-Lipschitz velocity fields may lead to interesting and non-trivial features in terms of transport by such fields. Namely, generalized flows, constructed as limits of regularized standard flows, typically appear to lack desirable properties such as stability with respect to perturbations or

Markovianity in time so that the failure of Lipschitz continuity should be unsurprisingly lead to physically impossible solutions that have no serious metaphysical import, as, for instance, in Norton's causal skeptical anti-fundamentalist programme.

I argue that indeterminism of the Norton-type Lipschitz-indeterministic systems should perhaps be viewed as an artefact of certain (infinite) idealizations essential for the models, depriving the examples of much of their intended metaphysical import, as, for example, in Norton's antifundamentalist programme.

PART II

Undecidability and Unpredictability

in Mathematics and Physics

3.1 Introduction

In Part II of the thesis, after reviewing the incompleteness and undecidability results in formal logic and mathematics and briefly introducing the necessary formal apparatus, I demonstrate the algorithmic undecidability of a certain class of propositions about future contingent events in situations where one takes into account the *provocative nature* of prognoses, and discuss its *physical* and *philosophical significance* as it bears upon principal unpredictability of the behaviour of mechanistic systems.

Here I consider a model of a social (or, more generally, mechanical physical) system where prognoses are not merely passive forecasts of future happenings, but where they actively *provoke* the very events the prognoses are about; a case where the events would not happen at all, if we had not previously put forth this prognosis. Of special interest is a subclass of all such prognoses, so called *self-fulfilling prognoses*. In the case of self-fulfilling prognoses, the very fact of formulating, or putting forth, a prognosis about the state of the system at a certain time in future *initiates*, or *triggers*, a series of changes within the system, in such a way that at that future moment the system assumes exactly the state described in the prognosis.

The (mechanical physical) systems considered in this section are perceived as never-ending computational processes, characterized by *recursive*, or *computable*, dynamics: the responses of the system to a prognosis for these models are assumed to be *fixed*, *known*, and *computable* on a step-by-step basis. (More precisely, the response function of the system to a prognosis – the "law" governing the evolution of the system – is characterized by a *recursive* function.) I argue that even for such systems,

notwithstanding their simple mechanical and totally computable appearance, the class of effectively undecidable propositions which express the (classical) truth-values of the (self-fulfilling) prognoses; in general, is not empty.

In such situations, any Laplace's (or Popper's or Landauer's) demon having *all* the information about the world *now* will be unable to predict all the future; in order to answer certain questions about the future it needs to occasionally resort to, or to consult with, a demon of a higher order in the computational hierarchy whose computational powers are beyond that of any Turing machine – an *Oracle*.

Unlike more typical settings in which self-referentiality is taken to obtain between a symbol (a word, a sentence, a statement, a language, etc.) and its own *semantics*, *meaning*, or *interpretation*, I set up self-referentiality to hold between the descriptions of the states of the same system, as contained in the prognosis and in the resulting state of affairs, across temporal slices.

After introducing the result, I discuss its *physical* and *philosophical significance* as it bears upon in principle *unpredictability* of the behaviour of mechanistic systems. Here I discuss the various known attempts to translate algorithmic undecidability into physically meaningful language. These include the discussion of the recursive undecidability of the halting problem, the undecidability of the rule inference problem (expressed by the question, "given a specified class of laws, usually a class of recursive/computable functions, which one of these laws governs a particular system?"), both from extrinsic and intrinsic perspectives, as well as a cluster of complexity and randomness related issues as they present themselves in the workings of (weakly) chaotic systems.

There is yet another important source of undecidability which seems to be more naturally related to the provocative prognoses framework. This undecidability is due to so called computational complementarity which usually appears in complementarity games in the theory of finite automata. It has been argued that typical physical measurement and perception processes may exhibit features which resemble computational complementarity and diagonalization: while trying to read off the "true" value of an observable of a system, a measurement interacts with the system and thus inevitably changes its state. It is true both for quantum and classical systems with the major difference being that quantum theory postulates a lower bound on the transfer of action by Planck's constant \hbar . I conclude Part II with the conjecture that the provocative prognoses framework can be extended to allow parallel analogues in the settings quite different from the context of (social, or simple mechanistic) systems. In particular, it may be helpful in both elucidating complementarity issues in quantum physics and in allowing parallel results from the quantum level to be brought to the "macroscopic" philosophical level.

3.2 Undecidability in Formal Logic and Mathematics

3.2.1 Incompleteness in Formal Logic

Published in 1899, Hilbert's *Foundations of Geometry* is the first precise formulation of a formal axiomatic method as applied to Euclidean geometry. An axiomatic system of some discipline provides a systematization of the truths of this discipline, usually some branch of mathematics or science. Originally it was hoped that such systematizations would compress all truths about the subject matter into a finite (or recursively

enumerable) set of axioms. The axioms were thought to contain the totality of all substantive information about the subject matter. Once you reach such a systematization, the rest of your work will consist in "merely" teasing out the logical consequences of the axioms. Logic is an instrument for unfolding all the information buried in the premises (axioms) into their logical consequences (theorems). As the study of the relations of logical consequence, that is, of relations of implication or entailment, logic is given an authority to carry out logical inferences, or to draw deductive conclusions. This is what is called the *deductive function* of logic (Hintikka 1996).

By 1930 research on Hilbert's programme of capturing all mathematics in a logical web was in its heyday: In 1929 Presburger had shown that arithmetic without multiplication is decidable (Presburger 1930), and in 1931 Skolem did the same for arithmetic without addition and successor operation. Finitary consistency proofs had been given for some restricted but interesting fragments of arithmetic, for example, by Herbrand in 1931. There seemed good reason to think that a finitary consistency proof would be given for formalized arithmetic soon (Epstein and Carnielli 2000).

Yet, in 1931 Hilbert and his formalist programme received a dramatic and devastating blow. Kurt Gödel, a then unknown young mathematician from the University of Vienna, produced a completely unexpected result showing that Hilbert's formalist goal was unattainable (Gödel 1931). No formal axiomatic theory rich enough to include arithmetic (alternatively, no *mathematical* theory as *weak* as arithmetic) can ever be proved consistent. The best a logician can achieve is the knowledge that a system is *inconsistent*. Furthermore, no consistent axiomatic system rich enough to include arithmetic can be complete: there must exist mathematical statements expressed in the

symbols of the system that can neither be proved true nor false using the rules of the system.²²

Having fixed a particular formal system, and a formula φ defined in terms of the symbols in the system, one of four possibilities can be true of φ :

- (1) φ can be proved true in the system.
- (2) φ can be proved false in the system.
- (3) φ can be proved both true and false in the system.
- (4) φ can neither be proved true nor false in the system.

Apart from the obvious options (1) and (2), the possibilities (3) and (4) complicate matters. The result (3) would show that the system is inconsistent: if (3) holds, then the system is meaningless because it can be used to show that *any* statement made in the language of the system is true.²³ The possibility (4), if true, would show that the system is *incomplete*. Furthermore, adding new axioms to an incomplete system never cures the problem: while this may allow previously undecidable statements to be decided (just add them as new axioms, for example), it always generates some new undecidable propositions (Barrow 1990).

²² In 1936 Gerhard Gentzen showed that the consistency of first-order arithmetic is provable over the weaker base theory of primitive recursive arithmetic with the additional principle of quantifier free transfinite induction up to the ordinal ϵ_0 . His proof highlights one commonly missed aspect of Gödel's second incompleteness theorem which often is taken to say that the consistency of a theory can only be proved in a stronger theory. The theory obtained by adding quantifier free transfinite induction to primitive recursive arithmetic proves the consistency of first-order arithmetic but is not stronger than first-order arithmetic. In particular, it does not prove ordinary mathematical induction for all formulae, while first-order arithmetic does. The resulting theory is not weaker than first-order arithmetic either, since it can prove a number theoretical fact – the consistency of first-order arithmetic – that first-order arithmetic cannot. The two theories are simply incomparable.

²³ When asked by McTaggart to show that 'If twice 2 is 5, how can you show that I am the Pope?', Bertrand Russell replied at once: 'If twice 2 is 5, then 4 is 5, subtract 3; then $1 = 2$. But McTaggart and the Pope are 2; therefore McTaggart and the Pope are one!' (cited from Barrow 1990, p. 256)

It is worth noting that option (3) is quite distinct from option (4). As a matter of fact, incompleteness of a system *protects* the system from inconsistency, for if we can find just one statement in the language of our system that cannot be proven, then this guarantees that it cannot contain inconsistencies like (3) whose presence would render *all* statements true. Incompleteness of a system should be seen as a demonstration of the fact that the system's scope and content cannot be captured simply by axioms and rules of logic which define it.

By replacing "false" by "unprovable" in the classical liar paradox in the form "this statement is false"²⁴, Gödel famously obtained a tricky but mathematically meaningful sentence that expresses its own unprovability (as viewed from outside the system). Gödel himself was well aware of this analogy. In his 1931 paper (Gödel 1986, p. 149) he said (translation of Svozil 1993),

"The analogy of this argument with the Richard antinomy leaps to the eye. It is closely related to the "Liar" too; [footnote 14: Any epistemological antinomy could be used for a similar proof of the existence of undecidable propositions]..."

The consequence of a claim like "this statement is unprovable" can be summarized by the following alternative: (i) if, on the one hand, this statement were provable in a given formalism, then this fact would contradict the message of the statement (i.e., its unprovability), which, in turn, would render the whole formalism *unsound*; (ii) if, on the other hand, this statement would be unprovable in a given formalism, this would confirm

²⁴ A passage in St. Paul's epistle to Titus (1:12-13) refers to Epimenides, a Crete of Cnossus: "One of themselves, a prophet of their own, said, 'Cretans are always liars, evil beasts, lazy gluttons.'" (See also Anderson 1970).

the message of the statement. As a result, the formalism would be *incomplete* in the sense that there exist true statements which cannot be proven. There is no other consistent choice other than rejecting (i) and accepting incompleteness (ii).

Similarly, other metamathematical and logic paradoxes (Kleene 1952) have been used systematically to derive undecidability or incompleteness results. The general pattern is proof by *reductio ad absurdum*: first, a statement is assumed to be true; this statement yields absurd (inconsistent) consequences; the only consistent choice being its unprovability or nonexistence. Mostly, absurd consequences are constructed by techniques similar to Cantor's diagonalization method (Svozil 1993).

How can Gödel's incompleteness result be possible, and what are the "resources" that are responsible for it? There exist at least two features which are noteworthy: self-reference and universality – the possibility to express (but not necessarily to prove) certain facts about a formal theory within the theory itself and the use of the "absolute" notion of truth. The first two phenomena occur only in theories which are "rich enough" to allow coding of metastatements *within* the theories themselves. Theories which are "too weak", i.e., theories in which metastatements cannot be coded within the theories themselves, do not feature incompleteness of this kind, although they are incomplete in a more basic sense. The next two sections will concentrate on these issues.

3.2.2 Undefinability of Truth

It is Tarski's famous impossibility result (1932) that classical truth cannot be defined in any formal theory rich enough to include arithmetic. It is closely related to Gödel's incompleteness result. Indeed, Gödel first arrived at his incompleteness result by

discovering the undefinability of arithmetical truth in a first-order language. Although not expressed in the original paper, Gödel himself considered the use of an "absolute" notion of truth to be the main feature of his incompleteness theorems. In his reply to a letter by A. W. Burks he writes (von Neumann 1966, p. 55, Feferman 1984, p. 554):

"I think the theorem of mine which von Neumann refers to is not that on the existence of undecidable propositions or that on the lengths of proofs but rather the fact that a complete epistemological description of a language A cannot be given in the same language A , because the concept of truth of sentences of A cannot be defined in A . It is this theorem which is the true reason for the existence of undecidable propositions in the formal systems containing arithmetic. I did not, however, formulate it explicitly in my paper of 1931 but only in my Princeton lectures of 1934. The same theorem was proved by Tarski in his paper on the concept of truth."

Absolute (logical) truth is a transfinite concept; it cannot be defined by any finite description. Suppose, for instance, that a finite description of a "universal truth machine" (UTM) exists. For any given arbitrary statement as input, the universal truth machine is supposed to produce outputs TRUE or FALSE, depending on whether the statement is correct or incorrect, respectively. Consider a liar-type input statement, "the universal truth machine (UTM) with a finite description will not output that this statement is true". Not unexpectedly, the machine cannot produce TRUE or FALSE without running into a contradiction. Therefore, the machine cannot decide *all* questions, contradicting the assumption that the universal truth machine decides all questions. Yet, somebody from the *outside* (i.e., someone who is not part of this truth machine) sees that the above

statement is correct, or TRUE_1 , but this results in an *extrinsic* notion of truth which is *stronger* than the "portion of truth" available to the truth machine. By simply adding this statement to the description of the UTM one could produce a new machine, call it UTM_1 . However, by the same argument, the new machine would not be able to decide the input statement, "the machine UTM_1 with a finite description will not output that this statement is true". Yet, this latter statement can be seen to be TRUE_2 , etc., forcing a hierarchy of notions of truth *ad infinitum*.

By allowing restricted "degrees" or "strengths" of truth, one could resolve the paradox of the liar or similar paradoxes by basically blocking these paradoxes from being formulated in the first place. In this sense, Gödel's incompleteness results amount to a formal demonstration that the notion of truth is too comprehensive to be grasped by any finite mathematical model (Svozil 1993).

3.2.3 Universality and Self-Reference

Another essential and recurrent feature of all logical paradoxes yielding Gödel's incompleteness result is *universality* – an ability to express certain facts about a formal theory within the theory itself using the idea of *programs as data*. Each of such theories should be powerful enough to allow programs to be written which "understand" and manipulate other *programs* which are encoded as *data* in some reasonable way. For instance, in the λ -calculus, λ -terms act as both programs and data; in combinatory logic, combinatory symbols manipulate other combinatory symbols; each μ -recursive function having a number in a Gödel numbering can be used as input to other μ -recursive functions; and Turing machines can interpret their input strings as description of other Turing machines. This notion is close to the idea of *universal simulation* (thus the term

universality), in which a universal program or machine U takes an encoded description of another program or machine M and a string x as input and performs a step-by-step simulation of M on input x (Kozen 1997).

Universality of a formal theory allows for the possibility of *self-reference*. Consider, for instance, a statement similar to the classical liar paradox: "this statement is true". Unlike the liar's case, however, no contradiction arises if one assumes that this statement is true or this statement is false. Yet, its *meaning* remains unclear, because it does not make any difference whether one chooses one of the two cases. This can be seen as an example of a more general situation when a symbol (e.g., a word, a sentence, a statement, a language, etc.) refers to its own *semantics*, *meaning* or *interpretation*; more precisely, if a symbol refers to the relation between itself and the object it stands for. The existence of logical pathologies like the liar paradox can be seen as an example of the fact that there exist objects which cannot be named by any (formally defined) finite language. (The concept of "truth" discussed in the previous section is another example.)

Despite appearances, there seems to be nothing wrong with self-reference *per se*. It has been argued that self-reference, if applied properly, yields well-defined and meaningful statements of mathematical and physical significance (e.g., Martin 1970, Smullyan 1994). The known troubles in the form of inconsistencies typically occur by attempting some sort of *complete* self-interpretation, via some kind of diagonalization technique. Indeed, these inconsistencies, properly interpreted and coupled with self-reference and various diagonalization techniques, provide one of the most general and powerful tools for investigating undecidability (Svozil 1993).

3.2.4 Diagonalization and Reduction

There are two major techniques for showing that problems are undecidable: *diagonalization* and *reduction*. The diagonalization technique was first introduced by George Cantor at the end of the 19th century to show that there are fewer real algebraic numbers than real numbers (Cantor 1874). Since then diagonalization, in one form or another, has become the all-purpose and most powerful way of investigating undecidability.

To illustrate the diagonalization technique in action, I will briefly construct an algorithmically definable function which is not primitive recursive.

Assuming we can effectively enumerate all possible primitive recursive functions, there exists an effectively computable one-to-one function associating the natural numbers and the class of all primitive recursive functions. Let g_x be the x th function in this list. Now we define a diagonalization function h by the following expression:

$$h(x) = g_x(x) + 1.$$

Since the addition of one is an effectively computable operation and since $g_x(x)$ is itself an effectively computable function, h must be effectively computable as well.

Assume further that h is primitive recursive. Then it should appear in the list of all primitive recursive functions somewhere, say, at y th place:

$$h = g_y.$$

Now we combine the two expressions and look at h at y :

$$g_y(y) = h(y) = g_y(y) + 1.$$

The above contradiction shows that the class of all primitive recursive functions does not include all algorithmically definable functions.

(One possible way to avoid this contradiction while maintaining the functional enumeration is to take the functions g_i to be *partial* functions. This way, in the last expression, since $g_y(y)$ need not be defined for all y 's, the contradiction $g_y(y) = g_y(y) + 1$ need not arise. This justifies enumerating all partial recursive functions later in the section on undecidability of provocative prognoses.)

If, using a diagonalization of some sort, we have established that a problem A , such as, say, the Halting problem, is undecidable, we may show that another problem, B , is undecidable by *reducing* A to B . Intuitively, reduction means transformation of instances of one problem to instances of the other problem in such a way that these two problems appear equivalent with respect to decidability/undecidability. Although reduction does not provide us with an effective procedure of establishing decidability/undecidability, it can be used (if coupled with more direct methods like diagonalization applied to the first problem) to tell us that if there existed a decision procedure for B , then this procedure could be applied to the transformed instances of A to decide it as well. If, on the other hand, no such procedure for A existed, reduction would conclusively prove that no decision procedure for B exists either.

3.3 Classical Computation

In this section I briefly address the abstract notion of (classical) computability, Turing machines, and computational complexity, of which we only need a few basic properties.

The exposition for the most part is informal and covers only the case of functions and predicates in one variable. The omitted details and the extension of the definitions and results to functions and predicates in several variables can be found in any standard textbook in the theory of computation, as, for example, in Garey and Jonhson (1979), Boolos and Jeffrey (1989), and Kozen (1997).

3.3.1 Turing Machines

Informally, an *algorithm* is a set of instructions to follow; in using it, "we only need to carry out what is prescribed as if we were robots; neither understanding, nor cleverness, nor imagination, is required of us" (Kleene 1967). An algorithm transforms its *input* (initial data) into some *output* (result). (If computation never terminates for some inputs, we get no result.)

Inputs and outputs for many theoretical models of computational devices, such as Turing machines, are typically strings. A *string* is a finite sequence of symbols (characters, letters) taken from some finite *alphabet*. If A is a set of the symbols of some such alphabet, by A^* we will designate the set of all strings over the set A .

Following Kitaev *et al.* (2002) (see also Turing 1936, Adler 1974), by a *Turing Machine* (TM) we will understand a conceptual device consisting of the following components:

- a finite non-empty set of symbols S called the *alphabet*;
- an element $\# \in S$ called *blank symbol*;
- a subset $A \subset S$ called the *external alphabet*; we assume that the blank symbol does not belong to A ;

- a finite set Q whose elements are called (*internal*) *states* of the Turing machine;
- an *initial state* $q_0 \in Q$;
- a *transition function* defined as a partial function

$$\delta : Q \times S \rightarrow Q \times S \times \{-1, 0, 1\}.$$

(The term "*partial function*" means that the domain of δ is a subset of $Q \times S$. A function that is defined everywhere is called *total*.)

Any Turing machine represents an algorithm, and there are infinitely many Turing machines that represent a particular algorithm. The above described components can be taken to represent a computer *program*, rather than its "*hardware*". We now briefly describe the "hardware" such programs run on.

The Turing machine has three parts: a *tape* divided into the squares, or *cells*, a *scanner* with a read-write *head*, and a *control device* which is a finite-state automaton. Each cell can carry one (and only one) symbol from the machine alphabet S . The tape is assumed to be infinite to the right, and, at the beginning of the computation, all filled in by the blank symbols. As computation proceeds, the symbol in a cell may be erased and replaced by another symbol from the alphabet S . Therefore, the content of the tape is an infinite sequence $\sigma = s_0, s_1, \dots$, where each $s_i \in S$.

The scanner with a read-write head moves along the tape one cell at a time, scans the content of one cell currently under the scanner's head, and, possibly, replaces the scanned symbol by writing another symbol from the alphabet in its place.

The control device that determines the behaviour of a Turing machine works as follows. This device is capable of being in any one of the internal states from the set Q of all possible internal states. At each step of the computation, device's being in some particular state q with the symbol s_p being under the head determine the action performed by the control device as the next step: the value of the transition function, $\delta(q, s_p) = (q', s', \Delta p)$, specifies the new state q' , the new symbol s' , and the shift Δp in the position of the head that it has to undergo on the next step (with $\Delta p = -1$, for example, interpreted as the head moving one position to the left).

More formally, by the *configuration* of a Turing machine we will understand a triple $\langle \sigma; p; q \rangle$, where σ is an infinite sequence s_0, \dots, s_n, \dots of elements of S , p is a non-negative integer (position of the cell/head counted from the left to the right with the first cell on the left being assigned "0"), and $q \in Q$. At each step of the computation the Turing machine changes its configuration $\langle \sigma; p; q \rangle$ as follows:

- (a) it reads the symbol s_p ;
- (b) it computes the value of the transition function: $\delta(q, s_p) = (q', s', \Delta p)$, and, if undefined, the TM halts.
- (c) it writes the symbol s in the cell in the position p of the tape, moves the head by Δp cells, and changes its current state s_p to state s' . Thus, the new configuration of the TM is the triple $\langle s_0, \dots, s_{p-1}, s', s_{p+1}, \dots; p + \Delta p; q' \rangle$, and, if $p + \Delta p < 0$, the TM halts.

Inputs and outputs of the TM are strings of symbols over A . An input string α is written on the tape (which is, as we have said, initially all filled with blank symbols), starting from the left end of the tape to the right. At the beginning of the computation the head is positioned at the left end of the tape too, with the initial state of the control device being q_0 . In other words, the initial configuration of the TM is $\langle \alpha \#\#\dots; 0; q_0 \rangle$. Finally, as the computation proceeds, the configuration of the TM is transformed step by step according to the rules stated above, and we get the sequence of configurations of the TM:

$$\langle \alpha \#\#\dots; 0; q_0 \rangle, \langle \sigma_1; p_1; q_1 \rangle, \langle \sigma_2; p_2; q_2 \rangle, \dots$$

As we have said, this process terminates in two cases: if δ is undefined, or if the head bumps into the (left) boundary of the tape $p + \Delta p < 0$. After that, starting from the left end, we read the tape to the right until we reach some symbol that does not belong to A (e.g., the blank symbol "#"). The string before that symbol will be considered the output of the Turing machine.

If a Turing machine never terminates on a given input string x , it is said to *loop* on input x . A Turing machine that halts on all input strings is called *total* (Kozen 1997), or a *decider* (Sipser 2006).

3.3.2 Computability, Recursiveness, Decidability, and Semidecidability

Given the set A^* of all strings over A , a Turing machine is said to *compute* a partial function $\varphi_{TM} : A^* \rightarrow A^*$, if for input string α , the machine eventually terminates with output string $\varphi_{TM}(\alpha)$; the value $\varphi_{TM}(\alpha)$ is undefined if the computation never terminates.

Now, a partial function $f: A^* \rightarrow A^*$ is said to be *computable*, or *(partially) recursive*, if there exists a Turing machine TM such that $\varphi_{TM} = f$. If this obtains, the function f is said to be *computed by TM* .²⁵

A set of strings is said to be *(partially) recursive* if its characteristic function is (partially) recursive. A set of strings is called *recursively enumerable*, or *r.e.*, if there is a partial recursive function whose domain (co-range) is exactly this set, meaning that the function is defined at x if and only if x is a member of this set.²⁶

A (one-place) predicate (a property of strings) is said to be *decidable* if the set of all strings having property P is a recursive set. A property of strings is said to be *semidecidable* if the set of all strings having property P is an r.e. set.

Finally, a Turing machine is said to *work in time $T(n)$* if it performs at most $T(n)$ steps for any input of size n . Similarly, a Turing machine *works in space $s(n)$* if it visits at most $T(n)$ cells for any computation on inputs of size n .

3.3.3 The Church-Turing Thesis and Universal Turing Machines

Any Turing machine can be identified with an algorithm in the informal sense. The not so obvious converse statement is called the *Turing thesis*:

"Any algorithm can be realized by a Turing machine."

²⁵ There are several terms of referring to the property of *recursiveness* of a partial function which are used more or less interchangeably: "effective", "computable", "effectively computable", "recursively computable", "mechanically computable", or "algorithmically computable". This presupposes the truth of the Church-Turing thesis (see the next section) and all of them mean just *partial recursive functions*. (Some refer to recursiveness as "general recursiveness"; others reserve the phrase "general recursiveness" for total functions only and refer to the recursiveness of partial functions as "partial recursiveness". We will not adopt this practice.)

²⁶ Still another, equivalent, characterization of r.e. sets is due to Matiyasevich's theorem which states that every r.e. set is Diophantine (and *vice versa*) – see section 3.4.5.

It is closely connected with the *Church thesis* (Church 1936) that gives an alternative and equivalent characterization of computable functions in terms of (partial) recursive functions:

"Any algorithm corresponds to a partial recursive function".

Since there is a provable equivalence between the classes of (partial) recursive functions and functions computable by a Turing machine, these two theses are often united in one single Church-Turing thesis. The term "Church-Turing thesis" itself seems to have been first introduced by Kleene (1967, p. 232):²⁷

So Turing's and Church's theses are equivalent. We shall usually refer to them both as Church's thesis, or in connection with that one of its ... versions which deals with "Turing machines" as the Church-Turing thesis.

Since the early stages of formal computational theory in 1930's, a number of various characterizations of computable functions have been proposed and investigated: Turing machines (Turing 1936), Post systems (Post 1936, 1943), Church's lambda-calculus (Church 1933, Kleene 1935), combinatory logic (Schönfinkel 1924, Curry 1929), Gödel's theory of recursive functions, cellular automata (von Neumann 1966), Register machines (Jones and Matijasevic 1984, Chaitin 1987), Diophantine machines (Matiyasevich 1993), etc. All of these systems purport to embody the idea of effective computation in one form or another. Though representing very different models of computation working on different types of data, they all turned out to be equivalent to each other (see, e.g., Rogers 1967, Kleene 1952, 1967, Maltsev 1970, Shoenfield 1967,

²⁷ For various formulations of the thesis, its history and common misunderstandings, see, e.g., Jack Copeland's article "The Church-Turing Thesis" in the Stanford Encyclopedia of Philosophy, <http://plato.stanford.edu/entries/church-turing/>.

1972, Chaitin 1987, Matiyasevich 1993). The Turing machine formalism, as the one most closely resembling a modern (classical) computer, shall be taken as basic in this and following sections. Besides the version of a Turing machine introduced above many other custom variations are known: multitape, multidimensional tape, two-way infinite tapes, non-deterministic, probabilistic, etc. They all too turn out to be computationally equivalent in the sense that they can all simulate each other.

(In probabilistic Turing machines there is a caveat concerning the possibility of a non-computable intrinsic parameter ε built in the probability distribution that leads to rather strange computational outputs. We will return to this point in Part III when discussing the resources which are believed by some to enable quantum adiabatic computers with powers to break the Turing limit.)

Any Turing machine as a finite object can itself be encoded by a string over a fixed alphabet. If we fix a finite alphabet A , we can consider a *Universal Turing Machine UTM* defined as follows. Its input is a pair $(\ulcorner TM \urcorner, x)$, where $\ulcorner TM \urcorner$ is the encoding of a machine TM with external alphabet A , and x is a string over A ($x \in A^*$). The output of the machine TM is $\varphi_{TM}(x)$. So, *UTM* computes the function u defined as follows:

$$u(\ulcorner TM \urcorner, x) = \varphi_{TM}(x).$$

This function is *universal* for the class of computable functions of type $A^* \rightarrow A^*$ to the effect that for any computable function $f : A^* \rightarrow A^*$ there exists some Turing machine TM such that $u(\ulcorner TM \urcorner, x) = f(x)$ for all $x \in A^*$. (The equality here should be

understood as meaning that either both $u(\ulcorner TM \urcorner, x)$ and $f(x)$ are undefined, or they are defined and equal.)

The existence of a Universal Turing Machine *UTM* is a consequence of the Church-Turing thesis since the description of Turing machines was algorithmic. However, unlike the thesis itself, this is also a mathematical theorem: such a machine *UTM* can be constructed explicitly and proved to compute the function u . (More precisely, the existence of a Universal Turing machine for the class of Turing-computable functions is a mathematical theorem; the existence of a Universal Turing machine for the class of computable functions is a consequence of this theorem together with the Church-Turing thesis.)

More generally, a notion of a *Universal Computer*, or a *Universal Computing Agent* can be introduced as the class of computational conceptual devices or automata on which it is possible to implement all effectively computable functions.

3.3.4 Classical Complexity Class P

The computability of a function does not guarantee that one can compute this function in practice: to compute it, an algorithm may take too much space or time, thus the importance of the notion of *effective* algorithms.

The notion of an effective algorithm can be formalized in different ways, leading to different complexity classes. One of the most important is the class **P** of *polynomial* algorithms. The following is only a brief exposition of computational complexity classification, more details can be found, e.g., in Harel (1987), Garey and Johnson (1979), van Leeuwen (1990), and Papadimitriou (1994).

We say that a function F , defined on the set \mathcal{B}^* of binary strings over the set $\mathcal{B} = \{0,1\}$, is *computable in polynomial time* if there exists a Turing machine that computes it in time $T(n) = \text{poly}(n)$, where the notation $f(n) = \text{poly}(n)$ means that $f(n) \leq Cn^d$ for some constants C, d and for all sufficiently large n . If F is a predicate, we say that it is *decidable in polynomial time*.

The class of all functions computable in polynomial time, or all predicates decidable in polynomial time, is denoted by **P**. (There may be complexity classes defined only for predicates.)

3.3.5 Nondeterministic Turing Machines, Class NP, and NP-completeness

While several different definitions of another important class **NP** of problems can be given, we will use Nondeterministic Turing machines for this purpose. A *Nondeterministic Turing Machine* (NTM), resembles an ordinary deterministic machine, but can nondeterministically choose one of several actions possible in a given configuration. More formally, a transition function of an NTM is multivalued: for each pair (state, symbol) there is a set of possible actions. Each action has a form (new state, new symbol, shift). If the set of possible actions has cardinality at most 1 for each state-symbol combination, we get an ordinary Turing machine (Kitaev *et al.* 2002).

A *computational path* of an NTM is determined by a choice of one of the possible transitions at each step; different paths are possible for the same input.

Having this definition of a Nondeterministic Turing machine, we can define the class **NP** as follows. A predicate L belongs to the class **NP** if there exists an NTM M and

a polynomial $p(n)$ with integer coefficients such that the following two conditions are satisfied:

- (1) If $L(x) = 1$, then there exists a computational path that gives the answer "yes"/"accept" in time not exceeding $p(|x|)$, where $|x|$ stands for the length of the string x ;
- (2) If $L(x) = 0$, then there is no path with the above property.

The class **NP** of nondeterministic polynomial time problems can be shown to contain problems which can be *solved by a nondeterministic oracle* and *verified* in polynomial time. The function of an *oracle* – a magic conceptual computational device, or a "black box", which, upon being consulted, immediately supplies the correct answer to the given problem, and the workings of which remain hidden for us throughout computation – is basically to produce right guesses.

Now, the useful notion of reducibility allows us to verify that a given predicate is at least as difficult as some other predicate.

A predicate L_1 is *reducible* to a predicate L_2 (we write $L_1 \propto L_2$) if there exists a function $f \in P$ such that $L_1(x) = L_2(f(x))$ for any input string x .

Finally, a predicate $L \in \mathbf{NP}$ is **NP-complete** if any predicate in **NP** is reducible to it.

If some **NP**-complete predicate can be computed in time $T(n)$, then any **NP**-predicate can be computed in time $\text{poly}(n) + T(\text{poly}(n))$. So, if some **NP**-complete predicate belongs to **P**, then $\mathbf{P} = \mathbf{NP}$. In other words, if $\mathbf{P} \neq \mathbf{NP}$ (which is probably true), then no **NP**-complete predicate belongs to **P**.

If we measure running time "up to a polynomial", then we say that **NP**-complete predicates are the "most difficult" ones in **NP**.

The key result in computational complexity theory says that **NP**-complete predicates do exist. One of such predicates, called SATISFIABILITY, or $SAT(x)$, expresses the property of a propositional formula containing Boolean variables and logical connectives \sim , $\&$, and \vee , of being satisfiable, i.e., true for some values of the variables. By a theorem proved by Cook and Levin:

(1) $SAT \in \mathbf{NP}$;

(2) SAT is **NP**-complete.

As a corollary, if $SAT \in \mathbf{P}$, then $\mathbf{P} = \mathbf{NP}$. Or, in other words, if $\mathbf{P} \neq \mathbf{NP}$ (which, again, is probably true), then SAT cannot be solved polynomially in time. We will return to the satisfiability problem in Part III, when we will talk about certain *quantum* algorithm which are claimed to be able to solve it (or, more specifically, a 3-SAT problem) in polynomial time.

3.3.6 Probabilistic Turing Machines and Class BPP

A Probabilistic Turing Machine (PTM) is similar to a Nondeterministic one; the difference is that choice among activities is produced by coin tossing, not by guessing. More formally, some (state, symbol) pairs have two associated actions, and the choice between them is made probabilistically, according to some probability distribution on the set of all strings. Each instance of this choice is controlled by a random bit. Usually it is assumed that each random bit is 0 or 1 with probability $\frac{1}{2}$ and that different bits are independent.

Given an input string, a PTM generates a probability distribution on the set of all strings such that different values of the random bits lead to different computational outputs, and each possible output has a certain probability.

With a constant $0 < \varepsilon < \frac{1}{2}$ ("the admissible error possibility"), a predicate L is said to belong to the class **BPP** (**B**ounded-error, **P**robabilistic, **P**olynomial time) if there exist a PTM M and a polynomial $p(n)$ such that the machine M running on input string x always terminates after at most $p(|n|)$ steps, and the following two conditionals are met:

- (1) If $L(x) = 1$, then M gives the answer "yes"/"accept" with probability $\geq 1 - \varepsilon$;
- (2) If $L(x) = 0$, then M gives the answer "no"/"don't accept" with probability $\leq \varepsilon$.

It can be shown that the class **BPP** remains invariant with respect to variations of the admissible error possibility, ε , as long as it is in the interval $(0, \frac{1}{2})$ and is computable.

Probabilistic Turing machines (unlike Nondeterministic ones, which depend on imaginary oracles for guessing the computational path) can be considered as real computing devices. Physical processes like the Johnson-Nyquist noise generated by the thermal agitation of the charge carriers inside an electrical conductor in equilibrium or radioactive decay, the randomness of which is guaranteed by the very nature of quantum mechanics, are believed to provide real physical sources of random bits. A quantum computer (see Part III) is another model of computation that is inherently probabilistic.

One of the central questions of complexity theory is whether randomness *increases* computational power, i.e., the question of whether there is a problem which can be solved in polynomial time by a Probabilistic Turing machine but not a

deterministic Turing machine, or whether deterministic Turing machines can efficiently simulate all Probabilistic Turing machines with at most a polynomial slowdown. With respect to the latter question, the current belief is that it is indeed the case, implying that $P = BPP$. However, it should be noted that, in fact, while the class **BPP** associated with problems solvable by Probabilistic Turing machines remains the same if the admissible error possibility, ε , is replaced by, say, $1/3$, things will change essentially if this parameter is replaced by some *noncomputable real* p . It turns out that having a noncomputable intrinsic parameter ε built in the probability distribution leads to rather strange computational outputs, which are typically avoided in the standard textbook expositions (Kitaev *et al.* 2002). However, this is exactly the resource which is believed by some to enable quantum adiabatic evolution computers to break the Church-Turing thesis. We will return to this latter point in Part III.

3.3.7 Many-Faced Undecidability of the Halting Problem

One of the first problems proved undecidable was the so called Halting problem²⁸ (Turing's proof went to press in May 1936, whereas Church's proof of the undecidability of a problem in the λ -calculus had already been published in April 1936). The halting problem is closely related to the question of how a mechanistic system will evolve or what an algorithm or an automaton will output or what theorems are derivable for a mechanistic system. Informally, the halting problem is a decision problem associated with the question whether or not, given a description of a program or an algorithm A , and a finite input x , (Church's version) $A(x)$ will produce a specific output in finite time, or

²⁸ In none of his work did Turing use the word "halting" or "termination". Turing's biographer Andrew Hodges (1992) does not have the word "halting" or phrase "halting problem" in his index. The earliest known use of the phrase "halting problem" is in a proof by Davis (1958, pp. 70-71). See also Copeland (2004, p. 40).

(Turing's version) whether A will *terminate* or *halt* on x . (Church's version reduces to Turing's if the termination or halting condition is identified with the production of output y .)

The result known as the *recursive undecidability of the halting problem* (Turing 1936) states that there exists no Turing-computable function (partial recursive function) which decides the halting problem. Given a description of a program and a finite input, it is in general impossible to decide whether the program finishes running or will run forever, given that input. (Recalling the definition of semidecidability, the halting function can be shown to be *semidecidable* by Turing machines).

One consequence of the halting problem's undecidability is that there cannot be a general algorithm which decides whether any given statement about natural numbers is true or not. The reason for this is that the proposition which states that a certain algorithm will halt given a certain input can be converted into an equivalent statement about natural numbers. If we had an algorithm that could solve every statement about natural numbers, it could certainly solve this one; but that would determine whether the original program halts, which is impossible, since the halting problem is undecidable.

Another important consequence of the undecidability of the halting problem is Rice's theorem which states that the truth of *any* non-trivial statement about the function that is defined by an algorithm is undecidable. It is a very powerful theorem substantially generalizing Turing's results and subsuming many other undecidability results proven as special cases, and showing that undecidability is the rule, rather than the exception (Rice 1953, 1956).

Rice's Theorem: *Every nontrivial property of the r.e. sets is undecidable.*

Note that for Rice's theorem (also known as The Rice-Myhill-Shapiro theorem) to apply, the property must be, first, a property of *sets*, not of Turing machines (i.e., it must be true or false independent of the particular Turing machine chosen to represent the set), and, second, the property must be *nontrivial* – it should be neither universally true nor universally false, i.e., there must be at least one r.e. set which satisfies the property and at least one which does not. There are only two trivial properties, and they are both trivially decidable.

As an example, consider the following version of the halting problem – the *1-halting problem*: Take the (nontrivial) property of a partial function f if f is defined for argument 1. It is obviously nontrivial, since there are partial functions which are defined for 1 and partial functions which are undefined at 1. The *1-halting problem* is the problem of deciding whether there exists an algorithm which defines a function with this property, i.e., whether the algorithm halts on input 1. By Rice's theorem, such algorithm does not exist. (It is important to keep in mind that the theorem holds for the *function defined by the algorithm* and not the algorithm itself. It is, for example, quite possible to decide if an algorithm will halt within 100 steps, but this is not a statement about the function that is defined by the algorithm.)

Yet another important consequence of the general undecidability of the halting problem is the so called *Maximal Halting Time*, or *Busy Beaver, problem*, relating predictability of the behaviour of a mechanical system to the limits on the amounts of resources that a halting machine of a particular size can consume, in terms of either time or space. In his 1962 paper *On Non-Computable Functions*, Tibor Radó, a professor at Ohio State University, thought of a simple non-computable function besides the standard

halting problem for Turing machines. Given a fixed finite number of symbols and k states, select those Turing machines with k states which eventually halt when run with a blank tape. Among these programs find the maximum number of non-blank symbols left on the tape when they halt. (Alternatively, find the maximum number of time steps before halting.) This function of k – the *Busy Beaver function* – is well-defined but uncomputable.

The recursive undecidability of the halting problem and the undecidability of the Busy Beaver Problem are closely linked with an important class of the *complexity-based constraints* on predictability of the behaviour of computable systems for *finite-time* predications, and, connected with it, the idea of *randomicity*. The former constraints become especially important in cases of prediction by simulation for (weakly) chaotic systems in which a "speed-up" of the real-time computation processes is generally impossible. As for the link to randomicity, sometimes this position is formulated as a very radical metaphysical programme, stating that all instances of "genuine randomness" in physics may eventually turn out to be undecidable features of mechanistic systems (Svozil 1993). We will return to these points in the sections 3.6.3 and 3.6.4.

Finally, one more powerful unsolvability result is worth mentioning here – the so called *recursive unsolvability of the rule inference problem*, closely related to the classical problem of induction (Gold 1967, Li and Vitányi 1992, Angluin and Smith 1983, Adleman and Blum 1991, Svozil 1993). Assume we have two universal computers, UTM_1 and UTM_2 . Assume further that the second computer UTM_2 has been programmed with some algorithm or program p , and the knowledge of this algorithm is not given to the first computer UTM_1 . The task of UTM_1 , called the *rule inference*

problem, is to come up with the "law" or algorithm p by analysing output behaviour of UTM_2 . The following theorem (originally due to Gold 1967, pp. 470-473) states that this task cannot be performed by any effective computation:

Recursive Unsolvability of the Rule Inference Problem: *There exist total recursive functions which cannot be "guessed", or inferred, by any particular universal Turing machine.*

This result, important in the context of language learning, or in the context of inferring the laws governing evolution of the system, is just another "face" of the same halting problem and can be interpreted in terms of the recursive unsolvability of the halting problem: there is no recursive bound on the time the guesser UTM_1 has to wait in order to ensure that its guess is correct.

In the sections that follow I will describe another famous undecidable problem (or, more precisely, a family of problems) – Hilbert's Tenth Problem. Recently the physicist Tien Kieu suggested a quantum computing procedure that, presumably, could solve Hilbert's Tenth. If true, existence of such an algorithm compatible with modern physical theories would threaten the classical concept of computability and the Church-Turing thesis, and would require us to reconsider the limits of predictability and mathematical knowledge. (In Part III of the thesis I will critically address this proposal in detail, showing its failure to perform purported hypercomputation.)

3.4 Diophantine Equations and Hilbert's Tenth Problem

3.4.1 Introduction

In 1900 at the Second International Congress of Mathematicians, held that year in Paris, the German mathematician David Hilbert delivered his famous lecture entitled "Mathematische Probleme". In this paper he put forth a list of 23 unsolved problems, or, more precisely, 23 groups of related unsolved problems, that he saw as the greatest challenges for twentieth-century mathematics (Hilbert 1900). The problem number ten in this list, now referred to as Hilbert's tenth problem, was to find a "process" (what we now call a method or an algorithm) for deciding whether an arbitrary Diophantine equation has an integral solution:

10. DETERMINATION OF THE SOLVABILITY OF A DIOPHANTINE EQUATION

Given a Diophantine equation with any number of unknown quantities and with rational integral numerical coefficients: *To devise a process according to which it can be determined by a finite number of operations whether the equation is solvable in rational integers.*²⁹

A Diophantine equation is an equation of the form

$$D(x_1, \dots, x_n) = 0,$$

where D is a polynomial with integer coefficients.

²⁹ Translation from German was taken from Matiyasevich (1993) which relies on the English translation of Hilbert (1900).

These equations were named after a Greek mathematician Diophantus of Alexandria who was the first to investigate such equations and ask questions about their solvability in rational numbers.³⁰

Diophantine equations typically have several unknowns, so it is customary to distinguish the *degree of the equation with respect to a given unknown x_i* and the *(total) degree of the equation*, i.e., the maximum, over all the monomials constituting the polynomial D , of the sum of the degrees of the individual variables in such a monomial.

Below are typical examples of Diophantine equations, with x , y and z the (integer) unknowns, the other (integer) variables given:

- $ax + by = d$. *Bézout's identity*; an example of a *linear* Diophantine equation. It states that if a and b are integers with greatest common divisor d , then there exist (not necessarily unique) integers x and y (called *Bézout numbers* or *Bézout coefficients*) satisfying the above equation.
- $x^n + y^n = z^n$. For $n = 2$ there are infinitely many solutions (x, y, z) , the *Pythagorean triples*. For example, the triple $(3, 4, 5)$ satisfies this equation as

³⁰ Almost everything known about Diophantus' life comes from a single 5th century Greek anthology, collecting a number of games and strategic puzzles, in the form of an epitaph that itself encodes a mathematical problem to be solved in integers: "This tomb holds Diophantus. Ah, what a marvel! And the tomb tells scientifically the measure of his life: God vouchsafed that he should be a boy for the sixth part of his life; when a twelfth was added, his cheeks acquired a beard; He kindled for him the light of marriage after a seventh, and in the fifth year after his marriage He granted him a son. Alas! Late-begotten and miserable child, when he had reached the measure of half his father's life, the chill grave took him. After consoling his grief by this science of numbers for four years, he reached the end of his life." The solution gives 84 years as the age at which Diophantus died. Different authors diverge as to when exactly it happened – we only know that he lived between 150 BC and 350 AD. There has even been divergence between writers as to the last syllable of his name. A consensus nowadays seems to hold that the name was Diophantos, not Diophantes. See (Heath 1964).

$3^2 + 4^2 = 5^2$. For larger values of n , $n > 2$, *Fermat's last theorem* states that no positive integer solutions x, y, z satisfying the above equation exist.

- $x^2 - ny^2 = 1$, where n is a nonsquare integer (*Pell's equation*). Studied by Brahmagupta in the 6th century AD and later by Fermat, it has infinitely many solutions that yield good rational approximations to the square root of the natural number n .

- $\frac{4}{n} = \frac{1}{x} + \frac{1}{y} + \frac{1}{z}$, or, in polynomial form, $4xyz = n(xy + xz + yz)$. The *Erdős–Straus conjecture* states that, for every positive integer $n > 1$, the rational number $4/n$ can be expressed as the sum of three unit fractions (an *Egyptian fraction* representation), with x, y , and z all positive integers.

In specifying a Diophantine equation it is essential to indicate the range of the unknowns. Hilbert, in his Tenth Problem, spoke of solutions in *rational integers*. Though much work in the theory of Diophantine equations has been done in the case of *algebraic integers*, this latter case will be completely ignored in the present exposition, and we shall use the term "*integers*" to refer to rational integers.³¹ Following the tradition in mathematical logic, we will also speak of the non-negative integers as the *natural numbers*; in particular, we shall consider 0 to be a natural number.

³¹ An *algebraic integer* is a number which is the root of an integer polynomial (i.e., an algebraic number) which, in turn, can be expressed with integer coefficients and leading coefficient 1 (a monic polynomial). This generalizes the distinction between an integer n (the root of $x - n = 0$) and a fraction a/b (the root of $bx - a = 0$). In more abstract terms, the ring of algebraic integers is the integral closure of the ring of integers in the field of algebraic numbers. A number x is an algebraic integer iff $\mathbb{Z}[x]$ is finitely generated as an abelian group (i.e., \mathbb{Z} -module). Examples of algebraic integers include the Gaussian integers and Eisenstein integers. See, e.g., Kleiner (1998), Conway and Guy (1996).

3.4.2 Hilbert's Tenth as a Decision Problem

By the time of Hilbert's address to the Congress, solutions for a large number of Diophantine equations had been found and many others had been proven to be unsolvable. Yet, for different classes of equations, or even for different individual equations, number-theorists had to invent different specific methods. What Hilbert asked in his problem "to devise" was a *universal method* for deciding solvability or unsolvability of an *arbitrary* Diophantine equation.³²

Hilbert's Tenth Problem is an example (and, out of all 23 of Hilbert's problems, the only example) of a *decision problem*, i.e., a problem consisting of (countably) infinitely many *individual subproblems* each of which requires an answer "Yes" or "No". In this case, each individual subproblem is specified by a particular Diophantine equation, and the expected answer of the method or algorithm would read either "Yes, the equation has a solution" or "No, the equation has no solution". The heart of the decision problem is the requirement to find a single *universal* method which could be applied to each of the individual subproblems that make up the bigger problem.

A *solvability proof* for a decision problem can be done either directly or indirectly. In the first case, one provides a procedure for finding the answer to any individual subproblem, while in the second case, one *reduces* the given decision problem to another, the solvability of which has already been proven.

³² Of course, in 1900 the development of the theory of recursive functions and algorithms was still about thirty years away and the concept of undecidability and unsolvability were not yet formalized. For the discussion whether this implies that the 10th problem was not a well-defined mathematical problem at the moment when it was stated see, for example, (Denef *et al.* 2000, pp. 2-3).

An *unsolvability proof* for a decision problem can also be either direct or indirect. In the latter case, one also reduces the original decision problem to another, but what is required here is a reduction in the *reverse* direction. Namely, to establish the unsolvability of a decision problem, one has to reduce *to it* another problem the unsolvability of which has already been proven. It is exactly by providing a chain of reductions of more and more complicated problems to Hilbert's Tenth Problem, Yuri Matiyasevich, a then unknown graduate student, in 1970 presented an elegant unsolvability proof for this problem and showed that two fundamental concepts arising in different areas of mathematics – the notion of *recursively enumerable* or *semidecidable set* of natural numbers from computability theory and the purely number-theoretic notion of *Diophantine set* – turn out to be equivalent (Matiyasevich 1970, 1993, see also Denef *et al.* 2000).

3.4.3 Systems of Diophantine Equations

Although Hilbert formulated his Tenth Problem in terms of finding a procedure for deciding whether any single arbitrary given Diophantine equation does or does not have a solution, Diophantus himself had considered *systems* of equations. Correspondingly, some textbooks give definitions of Diophantine equations as a system of polynomial equations with integer coefficients where the number of the variables in all equations exceeds the number of the equations in the system.

However, it is not hard to see that a positive solution of Hilbert's Tenth Problem would also provide a method for deciding solvability of a system of Diophantine equations. Indeed, consider the following system of k equations:

$$D_1(x_1, \dots, x_n) = 0$$

.....

$$D_k(x_1, \dots, x_n) = 0.$$

If (and only if) this system has a solution in integers (x_1, \dots, x_n) , this is a solution for the following single Diophantine equation

$$D_1^2(x_1, \dots, x_n) + \dots + D_k^2(x_1, \dots, x_n) = 0$$

has one. Moreover, the set of solutions of (2) coincides with the set of solutions of (3). Thus, for systems of Diophantine equations the number of equations is not such an essential characteristic as it is in the cases of linear algebraic or differential equations.

3.4.4 Solutions in Integers, Natural Numbers, and Rational Numbers

Consider the following Diophantine equation:

$$(x+1)^3 + (y+1)^3 + (z+1)^3 = 0.$$

If asked to find solutions in *integers* (as Hilbert did in his Tenth Problem), this equation, obviously, has infinitely many solutions – any triple of the form $x = -z - 2$, $y = -1$ would do. If a solution, on the other hand, is sought in non-negative x , y , and z , then the fact that the above equation has no solutions is not trivial at all. So, for a particular Diophantine equation, the problem of deciding whether it has an integer solution and the problem of deciding whether it has a non-negative solution are, in general, two separate problems. However, the decision problem of determining the existence or non-existence of non-negative solutions can be *reduced* to the decision

problem of determining the existence or non-existence of integer solutions, for example, in the following manner.

Consider an arbitrary Diophantine equation

$$D(x_1, \dots, x_n) = 0,$$

and suppose we are looking for its non-negative solutions. Now consider the following system:

$$D_1(x_1, \dots, x_n) = 0$$

$$x_1 = y_{1,1}^2 + y_{1,2}^2 + y_{1,3}^2 + y_{1,4}^2$$

.....

$$x_n = y_{n,1}^2 + y_{n,2}^2 + y_{n,3}^2 + y_{n,4}^2.$$

It is easy to see that any solution of this system in *arbitrary* integers includes a solution of $D(x_1, \dots, x_n) = 0$ in *non-negative* integers. Using the fact that every non-negative integer can be represented as the sum of the squares of four integers³³, the converse can be shown to be true as well: for any solution of $D(x_1, \dots, x_n) = 0$ in non-negative integers x_1, \dots, x_n , there are integer values of $y_{1,1}, \dots, y_{n,4}$ that yield a solution of this equation. Now, as was shown in section 3, the above *system* can be squeezed into a single equation

$$\tilde{D}(x_1, \dots, x_n, y_{1,1}, \dots, y_{n,4}) = 0,$$

³³ The Four Squares Theorem of Lagrange (1772) states that the Diophantine equation $x_1^2 + x_2^2 + x_3^2 + x_4^2 = a$ is solvable in x_1, x_2, x_3, x_4 for any non-negative a . See (Matiyasevich 1993, Appendix).

solvable in integers if and only if the original equation $D(x_1, \dots, x_n) = 0$ is solvable in non-negative integers.

While Hilbert in his Tenth Problem asked for solutions in integers, Diophantus himself sought solutions in rational numbers. Consider the following Diophantine equation:

$$D(x_1, \dots, x_n) = 0$$

whose variables x_1, \dots, x_n range over rational numbers. Let us introduce another Diophantine equation

$$\tilde{D}(r_1, \dots, r_n, q) = 0$$

with its variables ranging over integers, and defined by

$$\tilde{D}(r_1, \dots, r_n, q) = q^k D\left(\frac{r_1}{q}, \dots, \frac{r_n}{q}\right),$$

where k is the degree of the polynomial D . The polynomial \tilde{D} is homogeneous of degree k , and so equation $\tilde{D}(r_1, \dots, r_n, q) = 0$ certainly has the *trivial solution* $r_1 = \dots = r_n = q = 0$.

For this reason, the question whether a given homogeneous Diophantine equation is solvable or unsolvable is equivalent to the question of the existence or non-existence of its non-trivial solution. It turns out that the following two decisions can be shown equivalent: (1) the problem of determining the existence of a rational solution for arbitrary Diophantine equations and (2) the problem of determining the existence of a non-trivial integer solution for homogeneous Diophantine equations. Thus, a positive solution of the restriction of Hilbert's Tenth Problem to the case of homogeneous

equations would supply one with a method for determining the existence of rational solutions for arbitrary Diophantine equations that Diophantus was after.

3.4.5 Families of Diophantine Equations and Diophantine Sets

Apart from single Diophantine equations and systems of Diophantine equations, number theory also studies families of Diophantine equations. By a family of Diophantine equations, we understand a relation of the form

$$D(a_1, \dots, a_n, x_1, \dots, x_m) = 0,$$

where D is a polynomial with integer coefficients with respect to all the variables $a_1, \dots, a_n, x_1, \dots, x_m$, separated into *parameters* a_1, \dots, a_n and *unknowns* x_1, \dots, x_m . Fixing the values of the parameters results in the particular Diophantine equations that constitute the family. It should be noted that a family of Diophantine equations is *not* an infinite *system* of equations because the unknowns need not satisfy all the equations simultaneously, as would be the case for a system. Families of Diophantine equations are also known as *parametric equations*.

For different values of the parameters, one can obtain equations that do have solutions as well as equations that do not have solutions. Correspondingly, given a parametric Diophantine equation of the form $D(a_1, \dots, a_n, x_1, \dots, x_m) = 0$, we can define a set \mathcal{M} consisting of the n -tuples of values of the parameters a_1, \dots, a_n for which there are values of unknowns x_1, \dots, x_m satisfying this equation, i.e.

$$\langle a_1, \dots, a_n \rangle \in \mathcal{M} \Leftrightarrow \exists x_1, \dots, x_m [D(a_1, \dots, a_n, x_1, \dots, x_m) = 0].$$

The number n is called the *dimension* of the set \mathcal{M} , and the above equivalence is called a *Diophantine representation* of \mathcal{M} . Sets that have Diophantine representations are called *Diophantine sets*. Clearly, every Diophantine set has infinitely many Diophantine representations.

One may extend these concepts to functions, properties and relations. For example, a *function* is Diophantine if its graph is a Diophantine set. A *property* of natural numbers is called Diophantine if the set of all natural numbers having this property is Diophantine. Finally, a *relation* among n natural numbers is called Diophantine if the set of all n -tuples being in this relationship is Diophantine. Corresponding equivalences are called *Diophantine representations* of, respectively, the function, property or relation.

Below are some easy examples of Diophantine representations:

- the property of being an odd number can be represented by the following equation:

$$a - 2x - 1 = 0,$$

i.e., a is an odd number if and only if $\exists x[a - 2x - 1 = 0]$, with x ranging over the natural numbers;

- the relations of \neq , \leq and $<$ can be represented by the following equivalences:

$$a \neq b \Leftrightarrow \exists x[(a - b)^2 = x + 1],$$

$$a \leq b \Leftrightarrow \exists x[a + x = b],$$

$$a < b \Leftrightarrow \exists x[a + x + 1 = b];$$

- the relation of being relatively prime is represented by equation

$$ax_1 - bx_2 - 1 = 0;$$

- the property of being composite (non-prime) is represented by equation

$$a - (x_1 + 2)(x_2 + 2) = 0;$$

- the set of all positive integers which are not powers of 2 is represented by equation

$$a - (2x_1 + 3)x_2 = 0.$$

3.4.6 Undecidability of Hilbert's Tenth Problem

A condition, *necessary* for a set to be Diophantine, arises if we look at Diophantine sets from a recursion theory point of view. Given a family of Diophantine equations (a parametric Diophantine equation)

$$D(a_1, \dots, a_n, x_1, \dots, x_m) = 0,$$

it is possible to effectively enumerate, or list, all n -tuples from the Diophantine set M represented by this equation. Namely, having fixed some order over $(n + m)$ -tuples of possible values of all variables $a_1, \dots, a_n, x_1, \dots, x_m$, we only need to check for each successive $(n + m)$ -tuple whether the equality holds or not. If it does, we put the n -tuple $\langle a_1, \dots, a_n \rangle$ on the list of elements of M . In this way every n -tuple from M will sooner or later appear on the list, possibly with repetitions.

We recall that a *recursively enumerable* subset of \mathbf{Z}^k is, by definition, one that can be listed by an algorithm, possibly with repetitions, and a *recursive* subset of \mathbf{Z}^k is a subset for which there exists an algorithm which can test membership of an arbitrary k -

tuple in it. Thus, for a set M to be Diophantine it is *necessary* that M be effectively enumerable. Martin Davis (1953) conjectured that this condition is also *sufficient*.

MARTIN DAVIS' HYPOTHESIS

*The notions of **Diophantine** set and **effectively enumerable** set coincide, i.e., a set is Diophantine if and only if it is effectively enumerable.*

This conjecture, if right, would immediately mean the Hilbert's Tenth Problem is undecidable because examples of sets that are effectively enumerable but not decidable were well known already in the 1930's. However, it was a rather bold conjecture at the time since there hadn't been any strong informal evidence in its favour, whereas there existed much informal evidence *against* it. As an example, Davis' hypothesis implied the existence of a particular polynomial P such that the equation

$$P(a, x_1, \dots, x_m) = 0$$

was solvable if and only if a were a prime number. Hilary Putnam (1960) noted that such an equation could be rewritten as follows:

$$a = (x_0 + 1)(1 - P^2(x_0, x_1, \dots, x_m)) - 1,$$

so that every solution of the former equation could be extended to a solution of the latter by putting

$$x_0 = a.$$

On the other hand, since a is non-negative, for any solution of the latter equation it follows that the product of the right-hand side should be positive, which is possible only if

$$P(x_0, x_1, \dots, x_m) = 0.$$

Thus Davis' hypothesis implies the existence of a particular polynomial such that the set of all its non-negative values is exactly the set of all prime numbers. This corollary was considered by many as an informal argument against Davis' hypothesis.

Yet, in 1970, building on work on exponential Diophantine sets by Martin Davis, Hilary Putnam and Julia Robinson, Yuri Matiyasevich, a then graduate student at the Leningrad Division of the Steklov Institute of Mathematics, presented a negative solution to Hilbert's Tenth Problem. In his doctoral thesis he proved (Matiyasevich 1970, Davis 1973, Denef *et al.* 2000) what is now called

DMPR-THEOREM (CHARACTERIZATION OF DIOPHANTINE SETS)

A subset of \mathbf{Z} is Diophantine if and only if it is recursively enumerable.

A consequence of Matiyasevich's result is that there is Diophantine set which is not recursive; that is, there is a Diophantine equation $P(x_1, \dots, x_n, y_1, \dots, y_m) = 0$ such that no algorithm whatsoever can detect the set of n -tuples $\langle x_1, \dots, x_n \rangle$ for which $P(x_1, \dots, x_n, y_1, \dots, y_m) = 0$ has a solution in terms of y 's.

An immediate corollary of the DMPR-theorem gives

COROLLARY (THE NEGATIVE ANSWER TO HILBERT'S TENTH PROBLEM)

There is no algorithm which, given an arbitrary Diophantine equation, can test it for possessing integer solutions.

Though this *negative* solution of Hilbert's Tenth Problem does constitute the formal *solution* of the problem with which Hilbert himself would probably be satisfied, there is another, related, question, the answer to which would most likely be "No": *Would Hilbert be satisfied with the statement of the problem itself if he knew it would be "solved" in this way?*

As Matiyasevich explains this point, Hilbert's lecture took two and a half hours but still this was not enough to present all the 23 problems, so that some of them, including the 10th, were not presented orally at all, but just appeared in the printed version of the lecture. In addition, the Tenth Problem occupies less space than any other problem in the lecture. In particular, there is no motivation for the 10th problem, and one can only guess why Hilbert asked about solutions in "*rational integers*" only (Denef *et al.* 2000, p. 18-19).

A plausible answer can be that Hilbert was an optimist who believed in the *positive* solution of the problem in integers. If there existed an algorithm for solving Diophantine equations in integers, such an algorithm would allow one to solve equations in rational numbers as well. Namely, consider the following equation:

$$D(\chi_1, \dots, \chi_m) = 0.$$

Solving this equation in rational numbers χ_1, \dots, χ_m is equivalent to solving equation

$$D\left(\frac{x_1 - y_1}{z + 1}, \dots, \frac{x_m - y_m}{z + 1}\right) = 0$$

in non-negative integers $x_1, \dots, x_m, y_1, \dots, y_m, z$. The last equation, in turn, is equivalent to the following homogeneous Diophantine equation:

$$(z+1)^d D\left(\frac{x_1-y_1}{z+1}, \dots, \frac{x_m-y_m}{z+1}\right) = 0$$

where d is the degree of D .

Though reduction in the converse direction of solving Diophantine equations in rational numbers to solving homogeneous Diophantine equations in integers is less evident, it can proceed in the following manner.

One first transforms the original equation into

$$D\left(\frac{x_1}{z}, \dots, \frac{x_m}{z}\right) = 0,$$

and then into

$$z^d D\left(\frac{x_1}{z}, \dots, \frac{x_m}{z}\right) = 0.$$

At this stage, however, an additional trick is needed to ensure that $z \neq 0$ (see, e.g., Matiyasevich 1993, and Smoryński 1991).

So, asking *explicitly* about solving Diophantine equations in integers, Hilbert asked *implicitly* about solving Diophantine equations in rational numbers. A *positive* solution of the Tenth Problem, as it stands in the lecture, would immediately give a *positive* solution to the corresponding problem about solving Diophantine equations in rational numbers.

However, a *negative* solution of the original formulation of the problem does *not* imply a *negative* solution for the problem of solving Diophantine equations in rational numbers. In fact, homogeneous Diophantine equations constitute a very special subclass of all Diophantine equations and it is quite possible that for this *narrower* class a corresponding algorithm exists.

So, Hilbert's Tenth Problem can be understood in two senses:

- *narrower sense*, i.e., literally as the problem was stated by Hilbert in his lecture;
- *broader sense*, which includes other problems the solution of which would easily follow from a positive solution of the 10th problem as it was stated in the lecture.

In the narrow sense the Tenth Problem is considered to have been solved but solving Diophantine equations in rational numbers still remains today an open problem, and the progress in this direction has been rather meagre (Denef *et al.* 2000, p. 19).

3.4.7 Diophantine Machines

One of the implications of the DMPR-theorem is that Diophantine equations can be treated as computing devices. Adleman and Manders (1976, 1978) introduced the notion of *Non-Deterministic Diophantine Machine* (NDDM), specified as follows (Denef *et al.* 2000).

Given a parametric Diophantine equation

$$D(a_1, \dots, a_n, x_1, \dots, x_m) = 0,$$

the corresponding NDDM works in the following manner: on input a_1, \dots, a_n it guesses the numbers x_1, \dots, x_m and checks whether or not they satisfy the equation; if the equality holds, the n -tuple $\langle a_1, \dots, a_n \rangle$ is accepted.

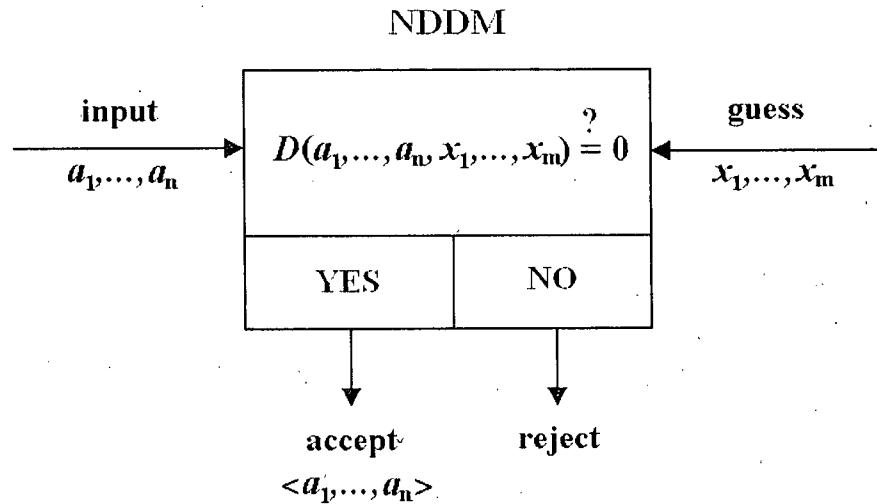


Fig. 22 Nondeterministic Diophantine Machine.

Notice that in the so introduced new computational device there is full separation of guessing and deterministic computation, with the computational part being very simple – the only calculation required is that of the value of a polynomial.

The DMPR-theorem implies that NDDMs are as powerful as any standard Turing Machines: every set recognizable by a Turing Machine is recognized by some NDDM, and, of course, *vice versa*. Adleman and Manders further supposed that, in addition, NDDMs are *as efficient as* Turing Machines.

Typically, one has to distinguish two different natural measures of computational complexity: TIME and SPACE. For NDDMs there is only one natural complexity measure which plays the role of both TIME and SPACE. This measure is SIZE, which is

understood as the size (in bits) of the smallest solution of the equation, defined either as the smallest possible value of $\max\{x_1, \dots, x_m\}$, or as the smallest possible value of the sum $x_1 + \dots + x_m$.

(Adleman and Mandres 1976, 1978) contain first results comparing the efficiency of NDDMs and conventional Turing Machines by estimating the SIZE of an NDDM simulating a Turing Machine with TIME in special ranges.

In particular, Adleman and Manders introduced the special class **D** of all sets **M** having Diophantine representations of the form

$$\langle a_1, \dots, a_n \rangle \in \mathcal{M} \Leftrightarrow \exists x_1, \dots, x_m \{ D(a_1, \dots, a_n, x_1, \dots, x_m) = 0 \ \& \ x_1 + \dots + x_m \leq 2^{|a_1| + \dots + |a_n|} \},$$

where $|a|$ denotes the (binary) length of a .

It can be easily shown that $\mathbf{D} \subset \mathbf{NP}$ since the class **D** is known to contain **NP**-complete problems. Adleman and Manders asked whether in fact the equivalence between the classes holds: $\mathbf{D} = \mathbf{NP}$.

3.5 Self-Fulfilling Prognoses and Laplacian Demon

3.5.1 Self-Reference and Self-Fulfilling Prognoses

Technically, the goal of this section is to demonstrate the algorithmic undecidability of a certain class of propositions about future contingent truths in situations where one takes into account the *provocative nature* of prognoses (Korolev 1998). Here I consider a model of a social (or, more generally, mechanical physical) system where prognoses are not merely passive forecasts of future happenings, but where they actively *provoke* the

very events the prognoses are about; a case where the events would not happen at all, if we had not previously put forth this prognosis. Of special interest is a subclass of all such prognoses, so called *self-fulfilling prognoses*. In the case of self-fulfilling prognoses, the very fact of formulating, or putting forth, a prognosis about the state of the system at a certain time in future *initiates*, or *triggers*, a series of changes within the system, in such a way that at that future moment the system assumes exactly the state described in the prognosis.

As a paradigm example of a self-fulfilling prognosis we can take a case of Girolamo Cardano. As the legend goes, he cast his own horoscope, and, having predicted that he would live only to the age of seventy-five, committed suicide on September 21, 1576, in order not to falsify his horoscope.³⁴

Within the social system context, this situation is possible due to the fact that prognoses may refer to the reality which itself is subject to human activity; in the meantime a choice of a particular activity is often determined by the pictures drawn by our expectations based upon the prognoses.

The systems I consider are assumed to be complex enough to allow self-fulfilling prognoses to occur. In addition, a mechanical physical system is perceived as a never-ending computational process, characterized by *computable*, or *recursive*, dynamics: the responses of the system to a prognosis for these models are assumed to be *fixed*, *known*, and *computable* on a step-by-step basis. (More precisely, the response function of the system to a prognosis – the "law" governing the evolution of the system – is

³⁴ For a more recent similar example see "Astrologer Misses Date With Death", Reuters, October 20, 2005, <http://www.smh.com.au/news/unusual-tales/astrologer-misses-date-with-death/2005/10/22/1129776001455.html>.

characterized by a *recursive* function.) I argue that even for such systems, notwithstanding their simple mechanical and totally computable appearance, the class of effectively undecidable propositions which express the (classical) truth-values of the (self-fulfilling) prognoses, in general, is not empty.

It turns out that the self-fulfilling prognoses are the bearers of principle algorithmic undecidability in the system. In general, having *all* the information about the world *now*, no Universal Turing machine predicting the future will halt before the moment when the event actually happens, and none of Laplace's (or Popper's or Landauer's) demons will be of any help here. Put differently, even if Laplace's demon knows all the initial conditions about the system with infinite accuracy, it requires, in addition, computational powers beyond that of any Turing machine, to be able to answer certain questions about the system's future. In computer science such power is attributed to a theoretical device called an *Oracle*.³⁵ So, Laplace's demon, to be able to do its intended job – predict the future – occasionally needs to resort to, or to consult with, a demon of a higher level in the computational hierarchy to be able to make such predictions. Behind any successful Laplace's demon necessarily there is another demon of a higher order, without which the subordinate demon would not be able to perform its function. If one accepts an outcome of a thought experiment that uses Laplace's demon, one is inadvertently accepting another demon lurking behind the first one; to bring the hiding demon to light would be to clearly see the computational powers required for the thought experiment to succeed.

³⁵ See Pitowsky (1996) for a similar result obtained via the detour through the generalized Bernoulli shift and Moore's (1990, 1991) realization of the Bernoulli shift.

(In Part III I will critically address a recent proposal, by Tien Kieu, to utilize a certain quantum adiabatic algorithm to serve as a *physical realization* of an *Oracle*, which, allegedly, is able to look through an infinite domain within a finite time. If actually implemented and coupled with a Laplacian demon (a Universal Turing machine), such a tandem would present a serious threat to traditional epistemological and ontological theories, thus requiring us to reconsider the age-old philosophical debates about the limits of predictability and mathematical knowledge.)

Historically, there have been several attempts to cash out the self-referential character of self-fulfilling prognoses in philosophically significant contexts. Thus, MacKey (1960) and ensuing discussion argued for the logical indeterminacy of a free choice, and Popper famously gave his arguments for the impossibility of historical prediction (and impossibility of an exact historical science, or historicism, with it) based on what he calls the "Oedipus effect" – the "influence of the prediction upon the predicted event, whether this influence tends to bring about the predicted event, or whether it tends to prevent it" (Popper 1961, 1950). Here I present a pure logical result in a very general context of forecasting the behaviour of mechanistic, totally computable systems, complex enough to allow self-fulfilling prognoses to take place.

3.5.2 Undecidability of Self-Fulfilling Prognoses

A system's "forecast" or "prediction" will be decomposed into two distinct stages or phases: the *algorithmic representation* or *description* of a system, and the actual *computation* of a prediction or actual *realization* of a prognosis, based on that representation. It will be assumed that the system under consideration is complex enough

to support universal computation (possibly up to limited computational resources).³⁶ Moreover, a social or mechanical physical system here shall be perceived as a never-ending computational process, characterized by computable, or recursive, dynamics. One appropriate *representation* or *description* of a computable system is an *algorithm* or *program code*, which, if implemented extrinsically on different computers, or intrinsically within the same system, yields a replica of that system. A mechanistic system can also be considered as a formal system, with the *index* or *Gödel number* of the axioms serving as description. In this sense the terms "(algorithmic) description", "program code", "index", or "Gödel numbers" are synonyms. Following a tradition in the theory of finite automata, we will interpret such descriptions as "natural" or "dynamical laws" governing the system (Svozil 1993).

Let p be some sentence (or a text) in an appropriate language L . The language L can be some formal language or a suitable chunk of some natural language such as English. The sentence p (or a text) is a *description* of some event e from the class E of all those events we are going to put forth prognoses about. We will assume that the language L and the sentence p are defined in such a way that having any such sentence we can always effectively determine whether or not it is a description of some event belonging to the class E . In other words, we assume the class DS of all sentences of the kind be an *effectively decidable* set.

³⁶ It is still an open problem to specify whether, for a given Newtonian dynamical system, there is a constructive embedding of a Universal Turing machine into its possible states of motion. For example, it is often conjectured that a system as simple as the three-body classical Newtonian system with point particles and gravity as the only force is already complex enough to be capable of universal computation (Pitowsky 1996).

This being so, there exists an effective one-to-one coding ν of this set by the natural numbers: $\nu : N \rightarrow DS$, where N is the set of the natural numbers. From now on we will identify any sentence p of DS with the code of this sentence at the coding ν . That is, instead of the sentence p we will talk of a natural number n such that $n = \nu^{-1}(p)$ or vice versa.

Accordingly, let us now, at t_0 , put forth a prognosis about an event taking place in the future, at some fixed time $t > t_0$, with a natural number x as a code of a sentence which is a description of the event.

Within this framework, any conceivable prognosis may be associated with some effective mapping $h : N \rightarrow \{0,1\}$ interpreted as follows:

$$h(x) = \begin{cases} 1 \leftrightarrow \nu(x) \text{ is true at } t \text{ (as seen at } t_0) \\ 0 \leftrightarrow \nu(x) \text{ is false at } t \text{ (as seen at } t_0). \end{cases}$$

Two things should be noted about prognoses as defined above. First, the qualification "as seen at t_0 " in the above definition is intended to emphasize the fact that, at the stage of pronouncing a prognosis now, at t_0 , we are assumed to be at freedom as to the prognosis' *content*, i.e., we are free to assign truth values to the future prognosed events as we wish. Whether or not some particular event *actually eventuates* (will, in fact, be true or false at t) does not and should not influence my choice of a particular prognosis at the stage of pronouncing, or putting forth, the prognosis. (It will be the system itself with its particular dynamics of responses to an external influence that will determine whether or not my prognosis can possibly come out true, i.e., be *in agreement* with what the prognosis states.)

Second, even though the mapping h is defined for all x 's to cover all possible, or conceivable, prognoses that can be formulated within the model, it is important to keep in mind that we are going to make a prognosis about only *one* event (at a time). The rationale behind this assumption is that the very fact of putting forth a prognosis about a particular event e_1 (say, about Baden-Württembergische Bank going bankrupt in three months) (or, to be more precise, people's *believing in*, or *taking seriously*, the prognosis) may *initiate*, or *trigger*, some changes, or processes, within the system (say, people's rushing to the bank to withdraw all their savings in the view of impending bankruptcy of the bank). Suppose these processes triggered by people's believing in the prognosis about e_1 's taking place at t eventuate in the system's assuming the state s_1 at t . Now, if, instead of the first prognosis, we were to put forth *another* prognosis, e_2 , this would initiate, or trigger, possibly *different* changes within the system, getting the system to evolve in a *different* way. Suppose the processes triggered by people's believing in prognosis about e_2 's taking place at t eventuate in the system's assuming the state s_2 at t . It is easy to imagine, however, that, if we were to put forth the two prognoses *simultaneously*, the processes triggered by people's believing in these two prognoses (i.e., about the compound event e_1 & e_2 taking place at t) could bring the system to a totally different outcome (different from s_1 & s_2), due to the possible *interference*, or *mutual influence*, of the processes on each other.

If seen on the level of *events* (those present both in a formulation of a prognosis and in the resulting state of the system), the latter possibility may presumably give rise to *non-Boolean* structures of events in the system whose dynamics is characterized by totally computable laws. Though this framework may be highly relevant when studying

complementarity in non-classical structures, we will not pursue this line of thought here (I will return to this topic again in section 3.5.6 on modeling classical and quantum measurements).

So, from now on we will identify the prognoses with the functions of the above sort.

Let us designate a class of all effective mappings from N to $\{0,1\}$ as H . We recall that a representation of a function, i.e., a syntactic expression for which there exist interpretations of these expressions as functions, if effective, must allow determination of an *algorithm* (a *program*, or a *description*) computing this function. We also note that, a prognosis has been put forth, the choice of a particular activity depends on the prognosis' content not directly, but via its descriptions (programs) that correspond to the (various) ways of actually realizing, or computing, this function. The same effectively computable function from H has a denumerable number of programs (descriptions) realizing this function.

Now we are ready for the formal definition of a *formulation of a prognosis*. Some Gödel numbering of all partial recursive functions (of one variable) has been fixed, we designate as φ_n a function that has its index n in this numbering. Then, by a *formulation of a prognosis* h we will understand any natural number n such that for all x the following identity holds:

$$\varphi_n(x) = h(x).$$

It should be noted that by no means *any given* natural number m is a formulation of some prognosis. In fact, the class M of all possible formulations of prognoses

$$M = \{n \in N : \varphi_n = h, h \in H\}$$

constitutes an *ineffective* subset of N .

An important qualification about the formulations of a prognosis is in order. Recall that any social or mechanical physical system within our model is perceived as a never-ending computational process, characterized by computable, or recursive, dynamics. We also assume that every step, or operation, in this computational process takes a fixed, arbitrarily small but *non-zero* amount of time. That is to say that any program φ_n , $n \in M$ for, or a possible way of computationally realizing of, our prognosis h necessarily takes some *finite* time before it halts (if it halts at all). If so, there may exist some programs for our prognosis made at t_0 about an event taking place at t that take time *exceeding* $T = (t - t_0)$. Even though such computational paths are perfectly admissible on *logical* grounds, we are going to exclude them from the admissible programs which allow computation of our prognosis within the time interval $T = (t - t_0)$ on *physical* grounds. (I will discuss the so called *complexity-based* physical constraints on predictability by simulation in more detail in section 3.6.3.)

Within the social system context, in addition to the complexity-based considerations, it will also be assumed that we will *not* take into account those prognoses which are written in a language that is not understood by a society – those knowingly will *not* affect upon the activity choice and are not essential for our discussion.

So, we will assume that the set of S of all *physically possible* formulations of a prognosis that we are going to take into account is an *effectively decidable* subset of the class M : $S \subseteq M$.

Let $u(x)$ be a function defined for all natural numbers such that if x coincides with a physically possible formulation of some prognosis (i.e., $x \in S$), then the function value is some certain activity α from the class A of all possible (at t_0) activities which we eventually choose if we believe in the prognosis with the formulation x .³⁷ That is, we assume that $u: N \rightarrow B$, where B is some class embracing the class A : $A \subseteq B$.

Further, let $r(x)$ be a function from B to N such that if $r(\beta) = n$ and $\beta \in A$, then n is the formulation of a prognosis which is *in agreement* with those events that will happen if we carry out the activity β .

Now let us introduce a function $f: N \rightarrow N$ taking it to be a composition of the two above functions: $f(x) = r \circ u(x)$. It is obvious that if x is a formulation from the set S , then the value of the function f will be a prognosis such that it will be in agreement with the outcomes of the activity which will be motivated by our believing in the prognosis with formulation x . We will work within a model where response of the society (a mechanistic system) to prognoses is characterized by a (*general*) *recursive function* f of the above sort.

Excluding the possibility of the prognoses which are knowingly *not* in agreement with what they state, we now go on to introduce a concept of a *permissible in f formulation of a prognosis*. By that we will understand such and only such a natural number n that

$$\varphi_n = \varphi_{f(n)}.$$

³⁷ Though quite an innocent assumption for mechanistic systems (our primary focus in this section), this would involve a serious simplification for social systems, possibly corresponding to the idea that the best (or most perfect) action (or the strategy) always exists and is unique.

The left-hand side of the equality can be seen as corresponding to the computational path realizing the prognosis in the case where no provocative character of the prognosis is involved, i.e., the case in which the response function of the system is taken to be just an identity function, $f(x) = x$. The right-hand side corresponds to the resulting computational path actually taken by the system after all the side-processes have been initiated, thus comprising all the provocative effects of the prognostication.

It is obvious that any natural number n satisfying the above identity is a formulation of a prognosis which, in the given society (system) with the given response function f , is knowingly in agreement with what it states (i.e., a self-fulfilling prognosis). And, vice versa, any natural number n which is not a permissible in f formulation of a prognosis, is either not a formulation of a prognosis at all, or is a formulation of such a prognosis that, at a given response function f , is knowingly in disagreement with what it states.

It is natural to associate the first case with the prognosis being true, and the latter case with the prognosis being false (one may call it the *Naïve Theory of Future Contingent Truth*).

According to the Kleene theorem of recursion (Rogers 1967, p. 180), given any (general) recursive function f there exists an effective procedure that allows to find a natural number n (called a *fixed-point value* for the mapping f) such that

$$\varphi_n = \varphi_{f(n)}.$$

After a fixed-point n of the mapping has been found, the question whether the found n is a physically possible formulation of the prognosis depends only on the set S and can be solved algorithmically (S is an effectively decidable set).

However, the problem of whether any *arbitrary* natural number m is a permissible formulation, generally speaking, is an algorithmically *undecidable* problem. The reason is that the set of all fixed points of a given recursive function need *not* itself be recursive. Indeed, there exist such recursive functions whose sets of fixed points are not even recursively enumerable, for instance, $f(x) = c$, where c is a fixed natural number.

A question arises: what is a class G of all recursive functions whose sets of "fixed points" are recursive sets? Not giving an answer to the question (we just note that it is not empty – it contains, for instance, the function $f(x) = x$) we emphasize its importance. If the society response function belongs to the class G , then among all permissible in f formulations of prognoses we can find formulations of any given prognosis h from H , i.e., the admissibility requirement of formulations of prognoses does *not* place any constraints on possible *content* of possible prognoses. And, vice versa, if f does not belong to the class G , then, in general, there is *no* effective way to determine whether a given formulation is permissible in f or not, even if some of them *are*, in fact, permissible.

Let us put this in a different way. In this situation it will be in general impossible to determine in an effective way whether the corresponding prognosis is true or false (with "true" and "false" understood as defined above). Even if we restrict ourselves to consider fixed, known, and computable responses of a society to an outer influence, the corresponding sentences expressing the truth-values of the prognoses, in general, are

effectively undecidable sentences. Not even Laplace's demon, having *all* the information about the world now, will be of any help here; and no Universal Turing machine predicting the future will halt *before* the moment when the event actually happens.

3.6 Undecidability in Physics

3.6.1 Introduction

Despite wide-spread recognition of the incompleteness results in formal logic, discovery of some other mathematical assertions – the continuum hypothesis and the axiom of choice – which are independent of the most common form of axiomatic set theory – Zermelo-Fraenkel set theory, and later proofs that "almost all" true theorems are undecidable (Calude *et al.* 1994b, see also Rice's theorem above), many mathematicians and physicists still long persisted in the view that all "real" mathematical and physical problems *are* solvable. Fair enough, after seeing the way in which Gödel's theorems were originally established it is rather hard not to become sceptical about their relevance to more practical matters. Emanating from the formulation of a rather innocent artificial linguistic paradox, Gödelian sentences appear to look like mere artefacts of a specific theory explicitly constructed for a specific metamathematical purpose, and, as such, seem to be only of pure *theoretical* interest. Far more impressive would be the demonstration that some great unsolved problems which have long tortured mathematicians and physicists *are* actually undecidable.

Gödel himself did not believe that the incompleteness results had any practical or physical significance, in particular for quantum mechanics.³⁸ However, with the rise of the theory of computation, recursion theory and the theory of finite automata, where undecidability refers not to statements being neither provable nor refutable in a formal system, but applies to *decision problems* taken as (countably) infinite sets of yes/no questions, new and more readily interpretable sources of algorithmic undecidability came into play, and many other sides of undecidability showed their faces.

The impact of the undecidability results in formal logic and mathematics on physical sciences has not yet been fully understood. Even though several attempts have been made to translate mathematical undecidability into a physically meaningful context (Kanter 1990, Komar 1964, Peres and Zurek 1982, Moore 1990, Da Costa and Doria 1991a, 1991b), a good many issues, both technical and philosophical, remain unsettled today (Kreisel 1974, Pour-El and Richards 1981, Calude *et al.* 1994a, Svozil 1993, 1995). In this section I will sketch the various possible ways of translating algorithmic undecidability into physical language. The following physical applications and their philosophical significance will be discussed below.

- (1) The general *problem of forecasting* the behaviour of a mechanistic, i.e., totally computable, system, and *inferring the recursive laws* governing a mechanical system, both in *extrinsic* and *intrinsic* contexts. The absolute knowledge of the dynamical laws governing the evolution of the system will be assumed to be given here by some external "oracle". The general problem of forecast is linked

³⁸ A rather anecdotal attempt to discuss a connection between Gödel's incompleteness theorem and the Heisenberg uncertainty principle that took place at the Institute for Advanced Study in Princeton in 1979 between Kurt Gödel and John Archibald Wheeler is recorded in Svozil (1993, p. 112) and Bernstein (1991, pp. 140-141).

to the recursive unsolvability of the halting problem. The impossibility of inferring the recursive laws governing the system (a mechanistic analogue of the problem of induction) is linked to the more general problem of identifying and learning a language and is constrained by the recursive undecidability of the rule inference problem, also reducible to the halting problem.

- (2) The *problem of predicting the behaviour of weakly chaotic systems*. Such systems are of special interest to physicists and philosophers of physics since, with respect to computational resources, they appear to be the fastest/optimal simulations of themselves, in that it is impossible to simulate a chaotic system with fewer than its own resources. This feature gives rise to a complex of in principle complexity-based constraints on the workings of such systems. The kind of undecidability at play here can be shown to be equivalent to the recursive undecidability of the halting problem, e.g., through Moore's proposal of physical implementation of the Bernoulli shift – the simplest example of a (weakly) chaotic system.
- (3) The *problem of the connection of instances of randomness in physics with the undecidable features of mechanistic systems*. As a radical metaphysical thesis it states that "there is no randomness in physics but a constant confusion in terminology between randomness and undecidability. God does not play dice" (Svozil 1993).
- (4) The *problem of modeling classical (and quantum) measurements*. The kind of undecidability which is in play here is due to so called computational complementarity which usually appears in complementarity games in the theory

of finite automata. A typical physical measurement and perception processes may exhibit features which resemble computational complementarity and diagonalization: while trying to read off the "true" value of an observable of a system, a measurement interacts with the system and thus inevitably changes its state. It is true both for quantum and classical systems with the major difference being that quantum theory postulates a lower bound on the transfer of action by Planck's constant \hbar . One may even speculate that quantum theory is the only theory so far that implicitly employs this kind of complementarity (Svozil 1993).

3.6.2 The Problem of Forecast of a Mechanistic System

In trying to predict the future, it has been a long tradition to imagine a being, or conceptual computational device – a demon – that would be able to do so given the knowledge of the dynamical laws governing the system and the system's present conditions. A Laplacian demon, within this framework, would be a very powerful computational device (a Universal Turing machine) with potentially unlimited computational resources (time and space) capable of measuring the initial conditions of the given system (or even the whole world) with arbitrary accuracy (Pitowsky 1996). Yet, apart from many difficulties which are mainly *practical* in character (e.g., the ability to measure the initial conditions with arbitrary, or "potentially infinite", accuracy) (Earman 1986, Feinberg *et al.* 1992), the ability to predict the future is hampered by various in principle computational constraints, many of which can be traced to the general recursive undecidability of the halting problem, and are, in fact, equivalent to the

halting problem in the sense that if (and only if) one could solve the problem in question, one could solve the halting problem.

One of the important computational limitations constraining our ability to predict the future deals with the general undecidability of the problem of inferring the recursive dynamical "laws" governing the system, or, more precisely, their *algorithmic descriptions*. Mechanistic systems here are treated as "black boxes" on which experiments are allowed only via some kind of input/output terminal. (By definition, for systems which are not mechanistic, no reasonable concept of a "recursive law" can be given, though probabilistic laws may still apply.)

First, we can consider a computational model of the system from an *extrinsic* perspective, with an algorithm representing a computable system implemented extrinsically on different computers. An idealized external experimenter (a demon) examines arbitrarily many copies of the system and, based on these observations, attempts to construct an algorithmic description of the system. As a result of the algorithmic undecidability of the rule inference problem this is in general impossible. There exists no systematic, i.e., effectively computable, way of deriving an algorithmic description of the system's laws from the input/output analysis of an arbitrary mechanistic system (Svozil 1993).

As no effective procedure is in general available for identifying the recursive laws of a mechanistic system, these laws can be thought of as given by some (extrinsic to the system) *oracle*. Such an oracle will be assumed to provide an algorithm which computes in an effective way enumeration of the system's evolution. Still, in the latter situation,

some (physically meaningful) propositions about the future of a totally computable system will be undecidable.

The simplest class of such propositions involves questions pertaining to *all* the future (as opposed to questions about the state of the system at any *specific time*). Examples of such questions are: "Is the solar system *stable*?", "Is the motion of a given system, in a known initial state, *periodic*?", "Is this particle ever going to reach this region of space?" These are typical and meaningful questions asked by physicists. What makes them insolvable is that all such questions involve *quantification over an infinite domain* of time instants. Thus, the question on stability of a mechanical system can be paraphrased as follows: "Does there exist some distance D such that *for all times* t the maximal distance between the particles constituting the system does not exceed D ?" The question on periodicity of a system of n mechanical particles is: "Does there exist some time T such that *for all times* t , the functions describing the coordinates of the particles of the system satisfy the following equality:

$$x_i(t+T) = x_i(t), \text{ for all } i = 1, 2, \dots, n?$$

Indeed, according to Rice's theorem – a generalization of the Turing undecidability result – almost all (i.e., all except for the set of measure zero) conceivable questions about the *unbounded* future referring to the state of a particle at some *unspecified* time turn out to be computationally insolvable – no Universal Turing machine can decide them in any finite time. Therefore, to answer these questions, a Laplacian Demon needs powers that exceed those of any conceivable (classical) algorithm. In other words, it needs to consult an oracle – a conceptual computational device capable of looking through an infinite domain within finite time (Pitowsky 1996).

Whereas the previous examples involved reference to the state of the system at some *unspecified* time, the undecidability of the halting problem, as viewed from extrinsic perspective, has implications for *finite-time predictions* as well. As an example, consider predicting weather on a certain day in future on the basis of present conditions. There exist computer programs that can quite reliably calculate the weather on the day after tomorrow. However, they occasionally take more than 48 hours to run, in which case we would be hard pressed to call such a computation a *prediction* (Pitowsky 1996). This is the class of so called *complexity constraints* on finite-time *predictions by simulation*, which arise naturally in investigating weakly chaotic systems. We will return to this issue again in the next section.

While in the extrinsic setup the experimenter (a demon) is allowed to carry out a complete simulation of the observed system without altering the original system, this is not necessarily so in the case of the *intrinsic* setup. There, any model simulation of the system is necessarily *part of that system*. As it can be expected, in such a setup the matters with predictability can only worsen.

Following Svozil (1993), the question of a description within the same system or process can be related to the question of whether an experimenter can possess an intrinsic "description" or "blueprint" of itself, thus linking the issue to the possibility of *self-reproduction*. Here, as shown by von Neumann (1966), (at least) two senses of the question should be distinguished, depending on the way in which self-reproduction is actually realized.

In the so called *passive* mode of self-reproduction, the self-reproducing system contains within itself a passive description of itself and reads this description in such a

way that the description cannot interfere with the system's operations (von Neumann 1966, pp. 125-126). It appears that it is possible to write a program which includes its own description and, through this description, is able to reproduce itself. (The Kleene fixed point theorem can be interpreted as a proof of the principle existence of such "viruses".) John Von Neumann, a Hungarian-born American mathematician and physicist, in his *Theory of Self-Reproducing Automata* was one of the first to propose a formal cellular automaton model of a universal self-reproducing automaton. Since then the concept of self-replicating machines, or von Neumann machines³⁹, has been rigorously studied in various contexts, one of the most intriguing of them being space exploration. The idea was to send a self-replicating spacecraft (a von Neumann probe) to a neighbouring star-system, where it would seek out raw materials extracted from asteroids, moons, gas giants, etc., to create exact replicas of itself. These replicas would then be sent out to other star systems, repeating the process in an exponentially increasing pattern.⁴⁰

While, in the passive mode, the "blueprint program" (interpretable as the "algorithmic description" of the system in terms of "laws" and "system parameters") was assumed to be given by some sort of an oracle, and so defined passively, i.e., without self-examination, in the *active* mode, the self-reproducing system examines itself and thereby constructs a description of itself (von Neumann 1966, p. 126).

³⁹ Von Neumann himself called them "Universal Assemblers".

⁴⁰ A number of interesting arguments against the existence of extraterrestrial intelligence have been produced that would account for the Fermi paradox – the question of why, given the moderate rate of replication of von Neumann machines and the history of the Universe, we haven't *already* encountered extraterrestrial intelligence if it is common throughout the universe. See, for example, Tipler (1981), and Sagan and Newman's (1985) response to him.

It can be shown that if a self-reproducing machine is specified as a device consisting of elements which can be *analysed*, *identified*, and, after the analysis, *restored* to their previous state, then *self-reproduction by self-inspection* can indeed be made. The machine, for instance, can be divided into two distinct parts with each part containing, in one form or another, an analysing and a constructing element. Initially, the first part analyses the second, which is taken to be passive, and constructs a copy of it. Then the first part activates the active mode of the second part, and becomes passive. Finally, the second part analyses the first part and constructs a copy of it. In this way, the machine examines itself and obtains a complete description of its own original constituent structure (Laing 1977).

However, for a more general class of automata, such strategies of self-reproduction by self-inspection are not available. One of the situations where Laing's strategy fails is when some parts of the automaton feature so called *computational complementarity* – the feature that prevents a "diagnostic" act of measuring some parameters of (part of) the machine from being "non-destructive" – any attempt to probe the system necessarily initiates some changes within the system, *irreparably* destroying the original copy. Though this feature appears naturally in the context of the quantum theory, some classical automata can also exhibit this feature.⁴¹ I will say more about this in the section 3.6.5.

3.6.3 Weakly Chaotic Systems and Prediction by Simulation

Given the recursive "laws" governing a mechanistic system's evolution, a universal Turing machine can simulate the system by encoding the system as a program and

⁴¹ See, e.g., Moore's "Gedanken-Experiments on Sequential Machines", in Shannon and McCarthy (1956).

performing the computation of the system's evolution on a step-by-step basis. If the dynamical laws are computable, it is always possible to simulate a mechanistic physical system *completely* and *in finite time* in the sense that any entity of the system can be brought into one-to-one correspondence with its computational simulation. However, due to the recursive undecidability of the halting problem and undecidability of the maximum halting time problem, there exist systems for which *no* effective computation can *predict* its behaviour in any reasonable time. For such systems there are no "computational shortcuts", and no optimization or "speed-up" with respect to time which would allow one to talk about *prediction* is possible. This is the effect of so called "deterministic chaos" or "chaos theory" which recently drew a lot of attention from a wide range of disciplines.

The first-ever rigorous examination of deterministic chaotic effects is (arguably) attributed to Jacques Hadamard who, in 1898, did the first study of the behaviour of a type of dynamical billiards ("Hadamard's billiards"). This system considers a free particle gliding along a frictionless surface with constant negative curvature – the simplest compact Riemann surface resembling that of a donut with two holes. Hadamard was able to show that all trajectories of this system diverge exponentially from one another as one tries to calculate the system's future state (Hadamard 1898, see also Steiner 1994).

At around the same time, Henri Poincaré (1899), while studying the three-body problem, found that there exist certain bounded non-periodic orbits which never approach a fixed point. Though the first version of Poincaré's results contained a serious error (Diacu and Holmes 1996), the final published version contained many important

ideas which consequently led to the development of the theory of chaos. Later, this and similar phenomena drew special attention of mathematicians and physicists under the name of ergodic theory (see, e.g., Arnold and Avez 1968, Mackey 1974, Mañé 1987).

Another example of such deterministic chaotic system that we discussed earlier is forecasting the weather. Edward Lorenz, an American mathematician and meteorologist, who studied weather forecasting and developed weather simulation programs, was the first to recognize the chaotic nature of the phenomenon. As the story goes, in 1961, he ran his weather simulation program using a Royal McBee LGP-30 computer. Trying to see a sequence of data again and to save time, he started the simulation in the middle of its course. To do so, he wanted to enter a printout of the data which corresponded to the conditions in the middle of his simulation which he had obtained last time. To his surprise, the weather that the machine began to predict was totally different from the weather calculated before. Lorenz tracked this down to the computer printout. The printout rounded variables off to a 3-digit number, but the computer worked with 6-digit numbers. This difference was tiny and the consensus at the time would have been that it should have had practically no effect. However, Lorenz had found that small changes in initial conditions produced large changes in the long-term outcome thus making a long-term accurate prediction of weather practically impossible.

These phenomena can be seen as particular examples of a more general class of systems, usually characterized by *non-linear* dynamics and extreme sensitivity of a system's evolution to tiny disturbances to the initial conditions.⁴² Such systems may exhibit a special form of physical behaviour that renders prediction difficult or

⁴² It was Lorenz who coined the terms "butterfly effect" and "attractor".

practically impossible. For such systems, the required degree of accuracy with which one has to specify, or measure, the initial conditions and velocities becomes very large (more precisely, the rate of divergence of system trajectories in phase space, typically stated in terms of Kolmogorov entropy or Lyapunov exponents, grows exponentially). It has even been argued that a proper definition of the chaotic systems must be put in terms of their inherent unpredictability and uncomputability (Stone 1989, Ford 1981, 1983, 1988, 1989, see also Batterman 1993, Batterman and White 1996).

The non-linear dynamics of physical systems in its extreme form of deterministic chaos theory has given a further kick to physical "irrationality". The success of chaos theory has led to questioning the "runaway" reductionism and predictability that have characterized science ever since the time of Newton, and on which the quintessential linear quantum mechanics is most firmly based. The fact that strongly nonlinear systems can behave fundamentally differently from systems in which nonlinearities are introduced as perturbations makes such systems appear every bit as counterintuitive and perplexing as can quantum mechanics. It appears that physical chaotic systems – even those with totally computable descriptions (i.e., systems whose evolution and essential parameters are computable on a step-by-step basis) – feature a form of complexity-based undecidability that constrains our ability to predict their behaviour *by running any simulation*. In the scientific literature this form of undecidability in physics is often referred to as "weak physical chaos" to distinguish it from stronger forms of chaos which exhibit truly random behaviour (e.g., due to quantum fluctuations in the initial conditions).

In a series of papers, Christopher Moore (1990, 1991) proposed a model of a fully fledged universal Turing machine (with an infinite tape) realized as a simple *finite* physical system. Unlike other proposals where the computer size changes the metric properties of space-time (Pitowsky 1990, Earman and Norton 1993), this model, though using arbitrary close spatial points in computation, adds no special difficulty apart from the one which already exists – the transmission of arbitrary large number of bits in a small time interval (the "blue shift") (Pitowsky 1996). Moore's construction is a physical implementation of the simplest weakly chaotic system, named the Bernoulli shift, and it provides an important link between the general undecidability results for Turing machines and the workings of (weakly) chaotic systems. This construction, although assuming the validity of Newtonian mechanics, is also consistent with the special and general theories of relativity.

The *Bernoulli shift map* is a numeric transformation $D : [0,1] \rightarrow [0,1]$ defined by the following equation:

$$D(x) = \begin{cases} 2x, & \text{if } 0 \leq x < 1/2 \\ 2x - 1, & \text{if } 1/2 \leq x < 1 \end{cases} ,$$

or, equivalently,

$$D(x) = 2x \bmod 1 .$$

The graphical representation of the map looks like:

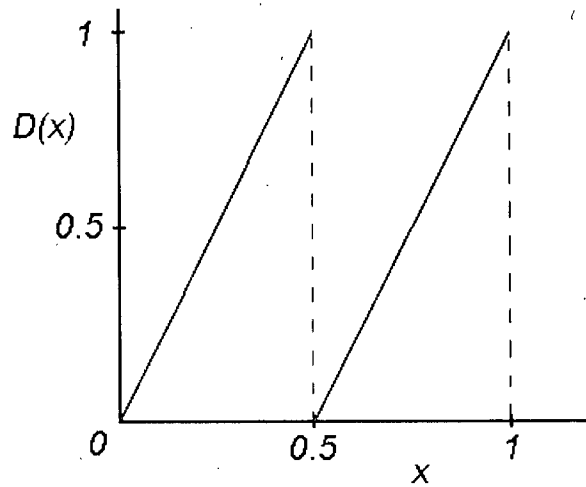


Fig. 23 The Bernoulli Shift Map.

The Bernoulli shift can be seen as an iterative system which starts with an initial parameter (the "seed" number), transforms this parameter, and then uses this transformed parameter for the input of the next iteration according to the formula:

$$x_{n+1} = 2x_n \bmod 1.$$

It can be shown that this system exhibits extreme sensitivity to initial conditions: if we initially start with two seed numbers separated by a tiny difference, the iterative application of the transformation makes the numbers diverge by about 0.5 (i.e., to fall on opposite sides of $x = 0.5$) very quickly, with the ratio of the final number difference to initial number difference growing exponentially with the number of iterations.

If the numbers appearing in the transformation are taken in their binary form, a seed number can be represented as follows:

$$0. a_0 a_1 a_2 a_3 a_4 a_5 \dots,$$

where the value of the a_i , $-\infty < i < \infty$, is either zero or one.

Multiplication by 2 in binary simply shifts all the numbers in the sequence to the left such that we get

$$a_0 . a_1 a_2 a_3 a_4 a_5 \dots$$

Taking the modulus drops the integer part, so we get

$$0 . a_1 a_2 a_3 a_4 a_5 \dots$$

More generally, the Bernoulli shift map can be defined as a transformation on an abstract sequence space of all *doubly* infinite sequences of zeros and ones:

$$\dots a_{-5} a_{-4} a_{-3} a_{-2} a_{-1} . a_0 a_1 a_2 a_3 a_4 a_5 \dots$$

The symbolic dynamics defined on this set of symbols runs on discrete time units (integers). After each time unit the system undergoes a shift to the left. If a_i is the number in the i th position at time t , then, at time $t+1$, the number in that position is $\sigma(a)_i = a_{i+1}$. The map σ is called the Bernoulli shift. The sequence before and after its shift thus looks like:

$$\dots a_{-5} a_{-4} a_{-3} a_{-2} a_{-1} . a_0 a_1 a_2 a_3 a_4 a_5 \dots,$$

$$\dots a_{-4} a_{-3} a_{-2} a_{-1} a_0 . a_1 a_2 a_3 a_4 a_5 a_6 \dots$$

The Bernoulli shift is a particularly simple example of a paradigmatic chaotic system that can be easily generalized to include more complex cases: If we use binary representation, it can be shown that, for sufficiently large k 's, the outcomes of the k th and the consecutive iterations from any given (irrational) seed number happen to be greater or less than $\frac{1}{2}$ as randomly as a sequence of heads and tails in a fair coin toss (e.g., no

matter how long a sequence of zeros and ones you measure, you cannot predict whether the next answer will be zero or one). Though the process is completely deterministic, the output of consecutive applications of the Bernoulli shift very quickly becomes indistinguishable from mere noise.⁴³

Moore (1990, 1991) showed how this model (in its more general form) can be physically (ideally) realized using a series of (ideal) parabolic and linear mirrors, and a beam of light moving between the mirrors. Just like its mathematical counterpart, the physical system realizing the Bernoulli shift is extremely sensitive to initial conditions – the initial positioning of a light particle before it starts bouncing back and forth between the mirrors: in order to predict the light particle position with accuracy of $\frac{1}{2}$ units, the experimenter (the demon) must measure the initial conditions with accuracy of $2^{-(k+1)}$, where $k+1$ is the number of rounds the light particle undergoes. The ratio of the final error to initial error is 2^{-k} , thus making it grow exponentially with time as measured by the number of rounds k .

Moore then demonstrated how any generalized Bernoulli shift can be put into one-to-one correspondence with some Turing machine (with a doubly infinite tape) which represents this shift, thus allowing transferral of the undecidability results from the usual theory of computation into the language of this system.⁴⁴

Thus, the undecidability of the halting problem, translated into the physical language of Moore's construction, says that there exists no algorithm which would predict whether the light particle ever enters the specific region in space which

⁴³ For more details, see, e.g., Schuster (1984).

⁴⁴ For more details refer to Moore (1990, 1991) and Pitowsky (1996).

corresponds to the halting state. This claim is valid even if we know the exact initial conditions with unbounded or infinite accuracy. Therefore, to answer the question: "Is the light particle ever going to reach this region of space?", Laplace's demon needs computational powers that exceed those of any possible (classical) algorithm. In other words, it needs to consult an *oracle*.

Finally, suppose that every step in this Universal Turing machine operation takes a fixed, arbitrarily small but non-zero amount of time. Suppose further that some demon tries to predict the state of Moore's system at time $t = 48$ hours. If Moore's construction simulates a very complex Turing machine there is no guarantee that the demon will accomplish the task on time – any demon's machine may be slower than the system itself. Indeed, a number of mathematical results, known as "speed-up" theorems (e.g., Enderton 1977, Pitowsky 2002) guarantee that the demon will fail occasionally in some such *finite-time* predictions.

This feature of weakly chaotic systems to be the fastest/optimal simulations of themselves to the effect that no computational "speed-up" is in general possible, can be illustrated with the following excerpt from Lewis Carroll's *Sylvie and Bruno Concluded*, (Chapter XI):

"We actually made a map of the country, on the scale of *a mile to the mile!*"

"Have you used it much?" I enquired.

"It has never been spread out, yet," said Mein Herr: "the farmers objected: they said it would cover the whole country, and shut out the sunlight! So we

now use the country itself, as its own map, and I assure you it does nearly as well."⁴⁵

The fact that in general it is impossible to simulate a chaotic system with fewer than its own resources raises an interesting question whether chaos in physics corresponds to the idea of randomness in mathematics, and, if so, whether randomness in physics (primarily in the quantum theory) merely reflects some undecidable features of mechanistic systems which are presently are not fully understood. We will return to this issue in the following sections.

3.6.4 Undecidability and Randomness

Another aspect of undecidability is its connection with the idea of randomness. This is another topic that has recently produced a lot of ado among mathematicians, physicists and philosophers of science, of which only a brief discussion is possible here.

Earman (1986, pp. 137-138) distinguishes two basic types of randomness which he calls, correspondingly, *process* and *product randomness*. Processes involving genuine chancy events as, for example, in the quantum theory, belong to the first category (also called *genesis randomness*). Processes involving outputs which lack any discernable pattern or which are 'out of shape' in one sense or another, exhibit *product randomness* (also called *performance randomness*). Surely, these two notions of randomness do not, in general, coincide. On the one hand, a sequence that is random in the product sense

⁴⁵ The ironic implication, of course, is that in truth it does almost as *badly*. A useful representation must be manageable, and this requires that it be *incomplete and inaccurate*. This situation is closely linked with the question of how the contents of a map – a *good* map, a *simplified* representation – are related to the real world. This problem (the *validation problem*) arises for models as well as for simulations, defined by Hartmann (2005b) as processes in which a model imitates the time evolution of a real system (Parker 2006).

need not necessarily be the output of a genuinely random process – any "random number generators" in (non-stochastic) computers or calculators producing "randomly looking" sequences of numbers are immediate examples that come to mind. On the other hand, genuinely stochastic processes may occasionally produce highly ordered sequences. It is possible (though very improbable), for example, that if we flip a (fair) coin 1,000 times, the coin turns heads all 1,000 times (Frigg 2005).

As it is the case with simplicity, it may nevertheless be very difficult to cash out our intuitions of what exactly the "lack of discernable pattern or rule" amounts to. Below is a very common way to define it rigorously.

Consider the following two sequences of numbers:

1) $\{7, 12, 5, 35, 22, 27, 3, \dots\}$ and

2) $\{1, 2, 4, 8, 16, 32, 64, \dots\}$,

or their binary representations (to be read by computers):

3) $\{111, 1100, 101, 100011, 10100, 11011, 11, \dots\}$ and

4) $\{1, 10, 100, 1000, 10000, 100000, \dots\}$.

To which extent and on which grounds can we say that the second sequence contains an easily discernable pattern while the first one lacks it (is *random*)?

To answer this question, we may consider the *length of the shortest* computer program that can generate each sequence. This length, in bits, can be taken to characterize the *complexity* of the sequence. If a sequence lacks any discernable pattern and contains no special rule for generating one entry from another, the shortest program can be nothing less than the sequence itself. If, on the other hand, the sequence exhibits

some pattern or order, then the program generating this sequence can be much shorter than the original sequence. In case of the second sequence the program will just list the powers of 2 (in decimal representation), or consecutively add one more "zero" to one "one" on the very left (in binary representation).

Correspondingly, we can define a sequence to be *random* if its complexity is equal to the length of the sequence itself. More formally, given a universal Turing machine, the *algorithmic complexity* of a sequence of symbols $\alpha_1\alpha_2\ldots\alpha_n$ is the length of the shortest program we have to provide in order to get this machine to reproduce (compute) this sequence. A sequence is then defined as *random* if the shortest program of this sort has the length of the sequence itself (i.e., the program basically says 'print $\alpha_1\alpha_2\ldots\alpha_n$ '). In this case it requires the maximum of information to specify the sequence (Chaitin 1987).

Using this notion of algorithmic complexity of a sequence of numbers, we can demonstrate how randomness as defined above is connected with Gödel's theorems. Consider a computer capable of working with the arithmetical symbols and operations. Suppose we give it the following instruction: "Print out a sequence whose complexity can be proved to exceed that of this program." Clearly, the computer cannot respond. Any sequence it generates must, by definition, have a complexity less than that of itself. A computer can only produce a numerical sequence that is less complex than its own program. It shows that there must exist undecidable statements. Having picked a particular sequence – call it R – whose complexity exceeds that of the computer system, the question "Is R a random sequence?" turns out to be undecidable for this computer system. The complexities of the statements " R is random" and " R is not random" are

both too great for them to be translated by the computer system. Neither can be proven nor disproved.

Among other things, this result places restrictions on the scope of any approach to the laws of nature on the basis of simplicity alone. The scientific analogue of the formalist programme in mathematics is the idea that, given any sequence of observations, we try to describe them by some mathematical law. Surely, there may be all sorts of possible laws capable of generating this particular data sequence – some simple, some very complicated and "unnatural". Scientists typically prefer to have the laws with the lowest complexity in the above defined sense that would most succinctly encode the information into a simple algorithm (Occam's razor). Yet, this approach will never allow us to *prove* that a particular law we have formulated is a complete description of nature – there will always exist undecidable statements framed in this language that can never be proven to be the most economical coding of the facts (Barrow 1990).

Another analogy that can be brought to the surface by these considerations is that between the intuitionist philosophy of mathematics and certain *epistemic* notions of determinism, as well as the connection of the latter with certain *ontological* views of determinism.

Intuitionists believe that the Platonic vision of the truth or falsity of (mathematical) statements as being independent of our knowledge of a procedure or algorithm which would decide the matter is indefensible. The only coherent concept of truth in mathematics, they argue, is ultimately *epistemic*, identifying truth with provability. While intuitionists typically do not require an actual human execution of a

complete proof (as long as there is an algorithm (a demon) which can, in principle, do it), a more extreme position, *finitism*, disagrees even about this contention.

When it comes to determinism, it has been a long tradition to draw the analogous line between the *ontological* and *epistemological* notions of determinism.

The ontological view of determinism takes the truth values of the statements about the future states of a physical system to be definite and (uniquely) determined by the state of the system *now*. These truth values reflect the feature of the universe to develop in a unique manner that is independent of our knowledge of it. The similarity of the ontological view of determinism with mathematical Platonism is clear: the Platonist takes the *set of all natural numbers* as an object with definite properties while the ontological view of determinism takes the *set of all future states of the universe* to be an object with definite properties (Pitowsky 1996).

Epistemic versions of determinism, on the other hand, tend to conflate determinism with predictability in principle – a physical system is deterministic just in case someone (preferably a demon with a reasonable extension of human capabilities) can predict the future state of the system from its present state.⁴⁶ The only thing left to be specified is what exactly constitutes this "reasonable extension of human capabilities". Although the notions of proof and construction are usually left by intuitionists to be infinitely extendable, in practice they identify constructive procedures with algorithmic procedures (compatible with intuitionistic inference rules). In words of Pitowsky (1996), the intuitionistic super-mathematician is an owner of a universal Turing machine.

⁴⁶ See, e.g., Earman (1986) attributing this view to Laplace.

Earman (1986) believes that any attempt to predicate determinism on predictability is a "confused and confusing brew" and recommends "that the notion of prediction with all its epistemological connotations be dropped altogether" from the definition of determinism, thus isolating the ontological issue regardless of whether or not any human agent or demon can know it.

However, if an epistemic view of determinism is the physical counterpart to intuitionism, then, from an intuitionistic (or a Laplacian) point of view, Earman's proposition contains circularity: the intuitionist challenges the idea of "independent existence" in mathematics exactly on the grounds that unless an existence claim is attached to a definite procedure to verify it, it is not clear what its reference is. Earman challenges the intuitionists with hopelessly confusing ontology and epistemology, thus assuming, without much of an argument, that one can make sense of the former independently of the latter. Yet, this is precisely what the intuitionist denies. In addition, since the physically significant mathematical structures employed in physics can be shown to be rich enough to allow a translation of many "negative" (in the sense of undecidability) results of the number theory into meaningful physical propositions, such a physical proposition can be taken as carrying a definite truth value just in case its mathematical counterpart does (Pitowsky 1996).

Returning to the idea of randomness as linked to a computing system's exhausting the available computational resources, there are additional considerations in support of Pitowsky's position. As modern science takes it, the fundamental physical laws are mathematical in nature. Any description, or prediction, of the behaviour of a physical system is carried out by application of mathematical operations or transformations.

Many physicists tacitly believe that these mathematical operations or transformations are implemented in some abstract mathematical spaces "inhabiting" the perfect immaterial Platonic realm. An alternative approach, represented most notably by John Archibald Wheeler (e.g., 1983, 1986) and Ralf Landauer (1967, 1986) stresses that real calculations involve real physical objects, such as computers, which are themselves subject to physical limitations, and take place within the real physical universe with its specific available resources. Just as any computer can be perceived as a physical system, any physical system may be perceived as a computational process. From this perspective, it is reasonable to investigate physical systems with concepts and methods developed by the computer science. This, in turn, opens the possibility of investigating various in principle computational constraints on the operations of physical laws. Landauer argues that these constraints are not merely a practical inconvenience, but constitute the very nature of physical law (Landauer 1967).

Recall Laplace's characterization of his calculating demon as an intellect "vast enough to submit the data to analysis". A demon living in an idealized Platonic world could indeed afford itself to be "vast enough". If, however, we adopt the Landauer-Wheeler view on the nature of physical law, any such demon would have to manage to do with the computational resources available in the real universe. Something that could not be calculated, even in principle, within the real universe cannot be regarded as a legitimate application of physical law. A Landauer's demon associated with recent cosmological models which place fundamental upper bounds on the informational content and information processing rate would necessarily inherit these limitations, and

thus will feature fundamental unpredictability of complex physical systems (Davies 1994).⁴⁷

Landauer's demon should be distinguished from Popper's demon. Although Popper (1982) requires that Laplace's demon should itself be part of the physical universe and bound by the laws of nature, he seems to see his demon as playing the role of a human super-mathematician or super-scientist, not specifying the roles and place of the physical laws within this universe. In particular, Popper's demon "must *not* be assumed to *ascertain initial conditions with absolute mathematical precision*" thus depriving it of his demonic nature, and leaving it only with the title (Pitowsky 1996). In this respect, Popper's view of (epistemic) determinism would rather be analogous to mathematical finitism, while Landauer's view would entangle the epistemic and ontological aspects of determinism on a more fundamental level. With the recent outrageous claims of Tien Kieu (the focus of Part III) that a concept of an *oracle* (a conceptual device capable of looking through an *infinite domain* within a *finite* time) can be actually physically implemented by using the resources of the quantum theory, the whole issue promises to be given a further non-trivial kick. Yet, as I argue in Part III, Kieu's claims are unfounded, and the quantum adiabatic hypercomputer fails to perform hypercomputation.

⁴⁷ Paul C. W. Davies (1994) gives a numerical value of the upper bound on predictability: the total number of bits available for computation within the known universe is of order 10^{120} ; the estimated threshold complexity of a physical system which would exhaust this computational resource is easily broken by any biological structure – a chain of about 60-90 aminoacids of 20 varieties (a peptide) allows for more three-dimensional conformations leading to different molecular structures than 10^{120} . Davies uses this result to argue that even simple biological systems do break this computational resource available to all laws of nature to work with ("causal openness of a system exceeding a certain threshold of complexity"), thus leaving a room for some emergent properties and laws at a higher level.

3.6.5 Modeling Classical and Quantum Measurements

In modeling classical or quantum measurements, the feature of so called complementarity may enter the scene and should be taken into account. While trying to read off the "true" value of an observable of a system, a measurement necessarily interacts with the system and thus inevitably changes its state. As an example, we may imagine a dark room with a ping-pong ball moving in it. Assume now that an experimenter, who is not allowed to "turn on the light", wants to find (measure) its position and velocity. Not being able to see, the experimenter, in order to do so, may try touching the ball. However, as the experimenter touches the ball, it changes its original position and velocity. This situation is typical both for quantum and classical systems with the major difference being that quantum theory postulates a lower bound on the transfer of action by Planck's constant \hbar (Svozil 1993).

If successive measurements of some features of a (mechanistic) system "disturb" each other through some kind of "interaction" of the system with the measuring device, i.e., the measurement of one observable makes it impossible to carry out the successive measurement of another observable, then we will say that this system exhibits *complementarity*. If a system features complementarity, the corresponding calculus of the propositions reflecting the experimental results will not in general be distributive and Boolean. The reason why classical physics does not involve any non-classical propositional calculus is that the "disturbances" induced by the measurement process can be made arbitrarily small, compensated, and thus eliminated from the overall account (Svozil 1986).

Returning to the self-fulfilling prognosis undecidability result, a similar logical provocative prognosis framework can be attempted to be applied to the setting of quantum measurements. In this regard I see a novel two-state-vector formalism of quantum mechanics, which has been particularly helpful for the analysis of experiments on pre- and post-selected ensembles (Aharonov and Vaidman 2001, Aharonov and Rohrlich 2005) as being a suitable setting for such an attempt. In this formalism, an experimental measurement can be thought of playing the role of a prognosis which initiates the changes within the system eventually guiding the observer to the predicted or expected outcome. I believe the self-fulfilling prognosis logical framework can be extended to help us understand better the nature of complementarity and elucidate the ways in which certain information about such systems gets lost, given the non-Boolean algebraic structures of these models (e.g., Pitowsky 1982, 1989, Demopoulos 2003).

PART III

Quantum Hypercomputation and Oracle Machines

4.1 Introduction

In Part III of the thesis I critically address a recent proposal, by the theoretical physicist Tien Kieu, to utilize a novel quantum adiabatic evolution algorithm to break through the Turing limit and perform *hypercomputation*. If true, such device could serve as a *physical realization* of an *Oracle* – a computational device capable of looking through an *infinite* domain within a *finite* time. If actually implemented and coupled with a Universal Turing machine, such a tandem would present a serious threat to traditional epistemological and ontological theories, requiring us to reconsider the age-old philosophical debates about the limits of predictability and mathematical knowledge. I argue that Kieu's claims are unfounded, and the quantum adiabatic "hypercomputer" fails to perform hypercomputation. Though quantum computers may indeed require redescription of the traditional problem *complexity* space, so far they retain the classical (recursion-theoretic) notion of *computability*. The delineation between complexity and computability may also be instructive in light of many recent remarks in the academic as well as in the popular literature in which quantum computers are depicted as all-powerful machines (Llyod 1995, Preskill 1998).⁴⁸

In section 4.2 I present a quantum algorithm that works on the principle of quantum adiabatic evolution. Originally, this algorithm was introduced by a group of physicists from the MIT to solve satisfiability problem in polynomial time. I also survey the currently known results regarding the quantum complexity classes and briefly

⁴⁸ The arguments in this part are the result of the collaborative work with Amit Hagar (Hagar and Korolev 2006, 2007a and 2007b).

discuss the possible resources responsible for the superiority of quantum computers over their classical counterparts.

In section 4.3, after reviewing the relevant theory of oracles, oracle machines, relative recursiveness and a hierarchy of degrees of unsolvability, I present the quantum adiabatic evolution "hypercomputer" designed to solve a paradigmatically unsolvable problem – Hilbert's Tenth problem. I then discuss the weaknesses of the proposal, pointing to its failure to perform the purported hypercomputation. Irrespectively of whether or not the class of physically realizable hypercomputers is non-empty, Kieu's quantum adiabatic algorithm is *not* the member of this distinguished club.

4.2 Quantum Adiabatic Computation

4.2.1 Introduction

Quantum computing brings together ideas from classical information theory, computer science, and quantum physics. In the past two decades it has evolved from a visionary idea (Feynman 1982) into one of the most lively and fashionable research areas (Nielsen and Chuang 2000). An explosion of interest in quantum computing was triggered by Peter Shor who, in 1994, presented the first quantum algorithm for fast factorization of large integers (Shor 1994). Shor's demonstration of how his algorithm could exponentially "speed-up" classical computation posed a serious threat to modern cryptography (and, with it, to home banking and any other information transfer via internet) which assumes that fast factorization algorithms do not exist. Since then tremendous progress in the field has been marked by the discovery of other fast algorithms (most notably Grover's algorithm for quantum database search (Grover

1996)), the invention of quantum key distribution, and most recently, popular press accounts of experimental successes in quantum teleportation, and the demonstration of actual three-, five-, and even seven-qubit quantum computers. According to one authority in the field (Aharonov 1998),

we now have strong theoretical evidence that quantum computers, if built, might be used as powerful computational tool, capable of performing tasks which seem intractable for classical computers.

Notwithstanding this excitement, and apart from the almost insurmountable problem of practically realizing and implementing a large scale quantum computer (Unruh 1995, Haroche and Raimond 1996), a crucial theoretical question remains open, namely – what physical resources are responsible for the putative power of quantum computing. Put another way, what are the essential features of quantum mechanics that allow one to solve problems or simulate certain systems far more efficiently than on a classical computer? Remarkable is also the fact that the relevance of features commonly thought of as essential to the superiority of quantum computers, e.g., entanglement and interference (Josza 1997), is recently being questioned (Linden and Popescu 1999, Biham 2004). Moreover, even if these features *do* play an essential role in the putative quantum "speed-up", it is far from clear *how* they do so (Fortnow 2003).

In this section, after reviewing the relevant known results in quantum complexity classes and briefly discussing the possible resources responsible for the superiority of quantum computers over their classical counterparts, I introduce the quantum adiabatic evolution algorithm of Farhi *et al.* Though this algorithm was originally designed for

speeding-up solving problems like satisfiability, Tien Kieu claimed that a similar scheme can be used to perform hypercomputation by being able to solve a problem equivalent to a paradigmatically unsolvable problem – the halting problem – in a finite time and using only finite resources.

4.2.2 The Class BQP and the Powers of Quantum Computers

This section surveys the currently known results regarding the quantum complexity classes and discusses the possible resources responsible for the superiority of quantum computers over their classical counterparts.

The class of decision problems that can be efficiently solved by quantum computers is called **BQP** ("Bounded error, Quantum, Polynomial time"). Since quantum computers run only randomized algorithms, the class **BQP** for quantum computers is the counterpart of **BPP** for classical computers. More formally, it can be defined as the class of decision problems solvable with a polynomial-time algorithm, whose probability of error is bounded away from 1/3 (Nielsen & Chuang 2000). A quantum computer is said to "solve" a problem if, for every instance, its answer will be correct with high probability. If that solution runs in polynomial time, then that problem is in **BQP**.

(Again, as it is the case with the class **BPP**, the choice of 1/3 other "bounded error" probabilistic classes the choice of 1/3 is not essential – changing this constant to any real (computable) number $p \in (0, 1/2)$ does not change the class **BQP**.)

BQP is suspected to be disjoint from **NP**-complete and a strict superset of **P**, but that is not known. Both integer factorization and discrete log are in **BQP**. Both of these problems are **NP** problems and both are suspected to be outside **BPP**, and hence outside

P. Both are suspected to not be **NP**-complete. There is a common misconception that quantum computers can solve **NP**-complete problems in polynomial time. Recall, however, that the classical complexity of integer factorization problem is not known, and it will not be totally surprising if a classical polynomial time algorithm for this task is found (this is an easier problem than the **NP**-complete problems). So, that quantum computers can solve **NP**-complete problems in polynomial time is still an open question, and, in fact, it is generally suspected to be false. In this respect Grover's search algorithm (Grover 1996) is better; though the speed-up is smaller (only quadratic), it is provable. Classical search does require (worst case) linear time in the size of the data.

Operators that appear in quantum computing devices are linear operators. Daniel Abrams and Seth Lloyd (1998) have shown that if a quantum computer could be designed with *nonlinear* operators, then it could solve **NP**-complete problems in polynomial time. However, as of today, there is no experimental evidence that such nonlinearities are present in nature.

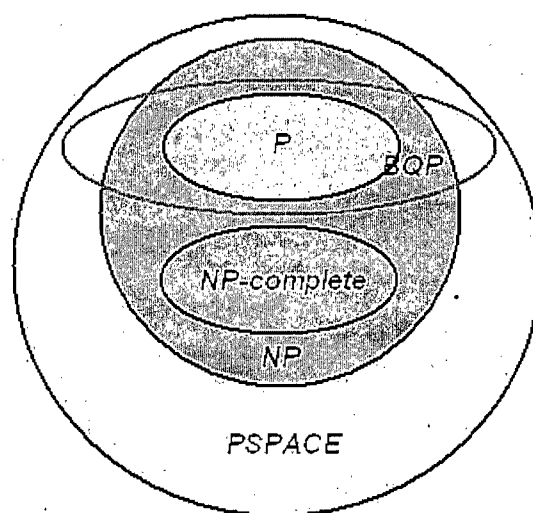


Fig. 24 The suspected relationship of BQP to other problem classes.

One of the embarrassments of quantum computing is the fact that, so far, essentially only one algorithm has been discovered, namely Shor's quantum Fourier transform, for which a quantum computer is significantly (exponentially?) faster than any known classical one. It is almost certain that one of the reasons for this scarcity of quantum algorithms is related to the lack of our understanding of what makes a quantum computer superior to classical computers. Quantum computer skeptics (Levin 2003) happily capitalize on this puzzle: if no one knows *why* quantum computers are superior to classical ones, how can one be sure that they *are*, indeed, superior?

When it comes to the possibility of quantum hypercomputation, several possible resources that might increase the powers of quantum computers over that of the Turing machine have been cited. For instance, accelerated Turing machines and infinite time Turing machines require that an *infinite amount of memory* be available for manipulations. Actual infinite memory required by these machines is to be contrasted with potential infinite memory (i.e, a finite, yet unbounded amount of memory) generally needed for a standard Universal Turing machine to function. Infinite storage capacities obviously present a serious *physical* obstacle for these proposals. Any attempt to store an infinite string of numbers while not occupying an infinite volume of space and not consuming an infinite amount of matter would have the disadvantage of *infinite precision* required to store and retrieve the appropriate bits (Hodges 2005). The infinite precision, however, is also problematic on the physical grounds: according to quantum mechanics, all physical quantities the continuous values of which one could attempt to use to represent an infinite string, seem to have principle limitations on how accurately they can be measured. For instance, we cannot measure distances more accurately than

the Plank length (10^{-35} metres). While some other, more exotic, means of implementing infinite memory storage have been proposed (such as infinite-dimensional Hilbert spaces with infinite quantum superpositions to allow *parallel* computations (Calude and Pavlov 2002)), it is still not at all clear that infinite memory is a physical possibility (Ord 2002).

Important in the Kieu's theoretical background is the role of genuine quantum *randomness* and, more specifically, that of *non-recursive information sources*. Recall that, one of the interesting questions of computation theory is whether randomness *increases* computational power, specifically, with respect to *computability*. Recall also that, while the class **BPP** associated with problems solvable by Probabilistic Turing machines can be shown to remain the same if the admissible error possibility, ε , that enters into the definition of a Probabilistic Turing machine, is replaced by, say, $1/3$, things change essentially if this parameter is replaced by some *non-computable real* p . Some proponents of the project of hypercomputation take this to be the key fact that leads to their goal. Yet, the question of whether by including some *non-computable* into a computational process can be *harnessed* in any way was already addressed by Shannon *et al.* (Shannon and McCarthy 1956). Their conclusion can be summarized thus: introducing a random element producing equiprobable outputs of 0 and 1 will, indeed, do no better than a deterministic Turing machine. If the probability (the admissible error possibility ε) is not $1/2$ but some other computable parameter, the situation is the same. If, however, the parameter ε is some *non-computable real number*, then the number ε can be computed. That is to say that indeed, computation is, formally speaking, extended to become computation relative to ε . However, contrary to the suggestions of Kieu who takes that to be the fundamental resource responsible for an enormous speed-up in his

scheme, it must be noted that the randomness in the output can do no better than merely *copying* the non-computable number that has already been put into the random element. In other words, it does not *create* anything non-computable and, therefore, seems highly implausible to be expected to infinitely speed-up classical computation (Hodges 2005).

4.2.3 Quantum Adiabatic Evolution Computer

4.2.3.1 Introduction

A well-known theorem of quantum mechanics – the quantum adiabatic theorem – was recently harnessed by a group of physicists from the MIT (Farhi *et al.* 2000a, 2000b) to develop a novel quantum algorithm. Their aim was to solve in polynomial time certain randomly generated hard instances of an **NP**-complete problem, and, in so doing, to provide another evidence that quantum computers (if large ones can be built) may outperform their classical counterparts.⁴⁹

The main idea behind the quantum adiabatic algorithm lies in the possibility of encoding a specific instance of a given decision problem in a certain Hamiltonian. One then starts the system in a certain (quantum) state corresponding to the lowest possible energy state of the system – the ground state – of another Hamiltonian which can be easily constructed (the initial Hamiltonian), and slowly evolves the system in time towards the interpolated and desired Hamiltonian (the final, or problem, Hamiltonian) which encodes the solution to the desired decision problem. Provided that certain conditions are met throughout the whole process of evolution, it can be guaranteed that the outcome of such a physical process will be the system in the ground state of the final

⁴⁹ It should be noted that it is still an open question whether the proposed quantum adiabatic algorithm does indeed yield an *exponential* speed-up. See, e.g., Reichardt (2004).

Hamiltonian, which, in turn, encodes the solution to the original decision problem. By physically measuring the ground state energy of the system at the end of computation one can get the solution to the original problem.

In the sections that follow, I present the quantum adiabatic algorithm of Farhi *et al.* which, presumably, can be used for solving a typical **NP**-complete problem – the satisfiability problem – and other combinatorial **NP**-complete problems in detail.

4.2.3.2 The Satisfiability Problem

Satisfiability is the problem of determining whether the variables of a given Boolean formula can be assigned classical truth-values in such a way as to make the whole formula true. Alternatively, it is the problem of determining that no such assignments exist, implying that the function expressed by the formula comes out false for all possible variable truth-value assignments. In the latter case, we say that the function (or corresponding Boolean formula of propositional calculus) is *unsatisfiable*; otherwise it is *satisfiable*.

The formal definition of the satisfiability problem (SAT) requires the function to be expressed in the so-called conjunctive normal form (CNF), i.e., as a conjunction of clauses where each clause is a disjunction of literals. For example, an n -bit instance of 3-SAT is a formula

$$C_1 \wedge C_2 \wedge \cdots \wedge C_M,$$

where each clause C_α involves (at most) three of the n bits:

$$C_\alpha = (z_{1\alpha} \vee z_{2\alpha} \vee z_{3\alpha}).$$

Each clause C_α is true or false depending on the truth values of some subset of the bits, and acts as a constraint on the possible truth values of its variables. For example, the clause

$$C = (z_1 \vee \sim z_2 \vee z_3)$$

is satisfied by all truth-value assignments to z_1 , z_2 , and z_3 except when z_1 and z_3 are false and z_2 is true.

An important result from complexity theory states that the class of satisfiable Boolean propositional formulas is **NP**-complete. In fact, 3-SAT was the first known **NP**-complete problem, as Stephen Cook demonstrated in (1971). Until that time, the concept of an **NP**-complete problem did not even exist. Since n -SAT (the general case) can be reduced to 3-SAT, and 3-SAT can be proven to be **NP**-complete, it can be used to prove that other problems are also **NP**-complete. This is usually done by showing how a solution to another problem could be used to solve 3-SAT. In practice it is typically easier to use reductions from 3-SAT than SAT to problems which one is attempting prove **NP**-complete.

In the context of quantum computation, many computationally interesting problems can be reformulated into an equivalent problem of finding a variable assignment that minimizes a so called *energy function*. In the case of 3-SAT problem, for instance, if each bit z_i ($0 < i < n$) takes the numeric value 0 or 1 (corresponding to the truth-values of true and false), and each clause C is associated with the 3 bits labelled i_C , j_C , and k_C , we can define an *energy function* for the clause C as follows:

$$E_C(z_{i_C}, z_{j_C}, z_{k_C}) = \begin{cases} 0, & \text{if } (z_{i_C}, z_{j_C}, z_{k_C}) \text{ satisfies clause } C, \\ 1, & \text{if } (z_{i_C}, z_{j_C}, z_{k_C}) \text{ violates clause } C. \end{cases}$$

Then we can define the *total energy function* E as the sum of the individual energy functions:

$$E_{TOTAL} = \sum_C E_C(z_{i_C}, z_{j_C}, z_{k_C}).$$

Obviously, $E_{TOTAL} \geq 0$, and $E_{TOTAL}(z_1, z_2, \dots, z_n) = 0$ if and only if (z_1, z_2, \dots, z_n) satisfies all of the clauses. Thus, finding the minimum energy configuration tells us if the formula has a satisfying assignment.

4.2.3.3 The Quantum Adiabatic Theorem

The possibility to encode a specific instance of a given decision problem in a certain Hamiltonian allows one to interpret the energy function of the problem as the *actual energy* of the physical system encoding this problem (hence the name of the function). At the end of computation, by *physically measuring* the ground energy level of the final Hamiltonian and checking it against the zero energy level one can determine whether the corresponding formula has a satisfying assignment.

More formally, suppose we can encode the solution to a given decision problem in the ground state of a Hamiltonian H_P acting in an n -qubit Hilbert space. This Hamiltonian (*Problem Hamiltonian*) is diagonal in the computational basis and is a member of a one-parameter family of Hamiltonians $\tilde{H}(s)$ varying smoothly for the parameter $0 \leq s \leq 1$. We then set a time-dependent Hamiltonian $H(t)$ as follows:

$$H(t) = \tilde{H}(t/T),$$

where $0 \leq t \leq T$, and T is the run-time of the algorithm. The Hamiltonian $H(t)$ governs the system state's evolution according to the Schrödinger equation:

$$i \frac{d}{dt} |\psi(t)\rangle = H(t) |\psi(t)\rangle,$$

whereas the run-time T determines how slowly H varies.

The Hamiltonian $H(t)$ has the following form:

$$H(t) = H_{C_1}(t) + H_{C_2}(t) + \cdots + H_{C_M}(t),$$

where each $H_{C_\alpha}(t)$ depends only on clause C_α and acts only on the bit in C_α . The initial state of the system, which is always the same and easy to construct, is the ground state of some fixed initial Hamiltonian $H(0)$. For each α , the ground state of $H_{C_\alpha}(T)$ encodes the satisfying assignments of clause C_α . The ground state of $H(T)$ encodes the satisfying assignment of the intersection of all the clauses. If the evolution time T of the computation is big enough, it can be guaranteed that the state of the system at time T (at the end of computation) will be very close to the ground state of $H(T)$, thus producing the desired solution.

The fact that a quantum system stays near its instantaneous ground state if the Hamiltonian that governs its evolution varies slowly enough is the content of the so called *quantum adiabatic theorem*. The rigorous exposition of the theorem (as, e.g., in Messiah 1961, pp. 739-746) considers a quantum system described in a Hilbert space \mathcal{H} by a smoothly time-dependent Hamiltonian, $H = H(t)$, for t ranging over $[t_0, t_1]$. Let

$U(t)$ be the time-evolution operator from time t_0 to $t \in [t_0, t_1]$. We denote $T = t_1 - t_0$. If the following conditions are satisfied:

1. For any $t \in [t_0, t_1]$, $H(t)$ has a purely discrete spectrum with eigenvalues denoted

$$E^1(t), E^2(t), \dots, E^i(t), \dots$$

We denote $P^1(t), P^2(t), \dots, P^i(t), \dots$, respectively, the projection operators on the eigenspaces.

2. The eigenvalues and the projectors are assumed to be continuous functions of t and there is no level crossing throughout the transition, i.e., the instantaneous eigenvalues remain distinct:

$$\forall t \in [t_0, t_1], E^i(t) \neq E^j(t) \text{ if } i \neq j.$$

3. $\frac{d}{dt}P^i, \frac{d^2}{dt^2}P^i$ exist and are bounded and piecewise continuous in the whole interval $[t_0, t_1]$.

Then, if the system is initially in the energy state

$$|\Phi_a^i\rangle \in P_a^i(t_0)\mathcal{H},$$

that is,

$$H(t_0)|\Phi_a^i\rangle = E^i(t_0)|\Phi_a^i\rangle,$$

then

$$\lim_{T \rightarrow \infty} U(t)|\Phi_a^i\rangle = P^i(t) \lim_{T \rightarrow \infty} U(t)|\Phi_a^i\rangle.$$

That is, for $T \rightarrow \infty$, if the system starts at t_0 in an eigenstate corresponding to the energy $E'(t_0)$ it will evolve (up to a phase) to an eigenstate corresponding to $E'(t)$ at t .

In the special case where $E'(t_0)$ is the ground state, the adiabatic theorem guarantees that in the limit $T \rightarrow \infty$ the system will remain in the ground state throughout its time-evolution.

Although in practice T is always finite, the more it satisfies a minimum "energy gap" condition, the less the system will deviate from the ground state. The *energy gap condition* states that there must exist a non-zero energy gap, g , between the ground state and the first excited state at *any* given time, and that $T \gg 1/g^2$. What governs the efficiency of the quantum adiabatic algorithm is thus the rate in which the energy gap between the ground state and the next excited state decreases with the increasing dimension of the Hamiltonian, i.e., with the size of the input.

4.2.3.4 The Problem Hamiltonian H_P

The transition from classical to quantum computation is typically done by replacing the classical bit z_i by a spin- $1/2$ qubit $|z_i\rangle$, where $z_i = 0, 1$. The states $|z_i\rangle$ are eigenstates of the i -th component of spin:

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

and

$$\frac{1}{2}(1 - \sigma_z^{(i)})|z_i\rangle = z_i|z_i\rangle, \text{ where } \sigma_z^{(i)} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The Hilbert space is spanned by the $N = 2^n$ basis vectors $|z_1\rangle|z_2\rangle\cdots|z_n\rangle$.

Clause C is now associated with the operator $H_{p,C}$:

$$H_{p,C}(|z_1\rangle|z_2\rangle\cdots|z_n\rangle) = h_C(z_{i_C}, z_{j_C}, z_{k_C})|z_1\rangle|z_2\rangle\cdots|z_n\rangle.$$

Finally, the Hamiltonian H_p associated with all the clauses is just the sum of Hamiltonians each acting on a fixed number of bits:

$$H_p = \sum_C H_{p,C}.$$

For so constructed Hamiltonian it can be shown that

- (1) $\langle\psi|H_p|\psi\rangle \geq 0$ for all $|\psi\rangle$ (the Hamiltonian is nonnegative), and
- (2) $H_p|\psi\rangle = 0$ if and only if $|\psi\rangle$ is a superposition of states of the form $|z_1\rangle|z_2\rangle\cdots|z_n\rangle$, where z_1, z_2, \dots, z_n satisfy all the clauses.

Thus, the ground state of the so constructed Hamiltonian encodes the solution to the original 3-SAT problem.

4.2.3.5 The Initial Hamiltonian H_I

Though the problem Hamiltonian H_p encoding the solution to a given 3-SAT problem is easy to construct, finding its ground state energy may not be an easy task. The idea behind the quantum adiabatic algorithm is to start the system with an easily constructed n -bit initial Hamiltonian, H_I , whose ground state is simple to find and construct.

Consider first the following 1-bit Hamiltonian $H_i^{(1)}$ acting on the i -th bit:

$$H_1^{(i)} = \frac{1}{2}(1 - \sigma_x^{(i)}), \quad \sigma_x^{(i)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

so that

$$H_1^{(i)} |x_i = x\rangle = x |x_i = x\rangle,$$

where

$$|x_i = 0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } |x_i = 1\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Now, for the clause C associated with the bits i_C, j_C , and k_C we construct

$$H_{1,C} = H_1^{(i_C)} + H_1^{(j_C)} + H_1^{(k_C)}.$$

Finally, we define the initial Hamiltonian as follows:

$$H_1 = \sum_C H_{1,C}$$

The ground state of H_1 , $|x_1 = 0\rangle |x_2 = 0\rangle \cdots |x_n = 0\rangle$, written in the z basis, is just a superposition of all basis vectors:

$$|x_1 = 0\rangle |x_2 = 0\rangle \cdots |x_n = 0\rangle = \frac{1}{\sqrt{2^n}} \sum_{z_1} \sum_{z_2} \cdots \sum_{z_n} |z_1\rangle |z_2\rangle \cdots |z_n\rangle,$$

which makes it easy to construct.

4.2.3.6 Adiabatic Evolution

We can then define the instantaneous eigenstates and eigenvalues of $H(t)$ by

$$H(t) |j; t\rangle = E_j(t) |j; t\rangle,$$

with the energy eigenvalues ordered as shown:

$$E_0(t) \leq E_1(t) \leq \dots \leq E_{N-1}(t),$$

where N is the dimension of the Hilbert space. Suppose $|\psi(0)\rangle$ is the ground state of $H(0)$, i.e.

$$|\psi(0)\rangle = |j=0; t=0\rangle.$$

According to the adiabatic theorem, if the gap g_{\min} between the two lowest levels, $E_1(t) - E_0(t)$, is strictly greater than zero for all $0 \leq t \leq T$, then

$$\lim_{T \rightarrow \infty} |\langle j=0; t=T | \psi(T) \rangle| = 1.$$

Having chosen and constructed the initial and the final Hamiltonians, we can define the interpolating Hamiltonian as follows:

$$\tilde{H}(s) = (1-s)H_1 + sH_p,$$

where $s = \frac{t}{T}$, so that

$$H(t) = \left(1 - \frac{t}{T}\right) H_1 + \frac{t}{T} H_p.$$

Under these conditions, and in the adiabatic limit, i.e., for T long enough, the evolution from $t = 0$ to $t = T$ (starting in the ground state of H_1) will lead to the ground state of H_p , that is, to the solution of the original computational problem.

4.3 Quantum Adiabatic Evolution Hypercomputation

4.3.1 Beyond Undecidability: Oracle Machines

We know that almost all interesting questions about workings of Turing machines – whether a given Turing machine halts on an empty tape, whether a given Turing halts on every input string, or whether two given Turing machines always produce the same output, etc. – are undecidable. But are all these questions equally hard? Suppose that we were given the power to decide the halting problem, by some magic. Could we somehow utilize this power to decide all undecidable problems? It certainly seems not unreasonable to ask whether a certain problem is decidable *relative to the halting problem*.

Questions about *relative computability* are typically formalized and studied using *Oracle Turing machines* – conceptual devices having special powers enabling them to perform certain computational tasks, which typically (but not necessarily) cannot be decided by usual Turing machines. Oracle machines were first defined by Turing in his Ph.D. thesis (supervised by Church). He described them as "new kind of machine" and called them "O-machines" (Turing 1939). Informally, an *Oracle Turing machine* is a usual Turing machine equipped with a black box – an *Oracle* – which is able to decide certain decision problems in a single step. The latter decision problems can belong to any complexity class, including known undecidable problems like the halting problem.

More formally, consider a procedure implemented on an ordinary Turing machine with a given finite set of instructions and input. Given some set X and an input, the entire computation proceeds algorithmically except that

- (a) from time to time, the computing agent may be required to answer a *membership question* of the form "is n in X ?", for a certain given set X (in general this question itself, i.e., the value of n , is the result of preceding calculation);
- (b) no means of answering such questions about X are given by the instructions of the Turing machine;
- (c) obtaining an answer to such a question counts as a single step in the overall procedure; and
- (d) subsequent steps in the procedure depend, in general, upon that answer.

If such answers are correctly and automatically supplied by some external agency, the computation is well-defined and effective. (If the set X itself is recursive, the entire procedure can be made recursive by adding instructions for computing the characteristic function of X .) Such a procedure is called an *algorithm relative to X* .

An external agency which supplies correct answers to questions about X in a finite time is called an *Oracle*. Sometimes it is pictured as an otherwise unspecified "black box" associated with the computing agent (Rogers 1967).

An alternative characterization of an *Oracle Turing machine* takes it to be a usual Turing machine which, in addition to its ordinary read/write tape, is equipped with a special one-way-infinite read-only input tape on which some infinite string is written. The extra tape is called the *auxiliary tape*, or the *Oracle tape*, and the string itself is called the *Oracle*. The machine can move its oracle tape head one cell in either direction in each step of computation and make decisions based on the symbols written on the oracle tape. Other than that, it behaves exactly like an ordinary Turing machine. If the

oracle is an infinite string over $\{0,1\}$, this string can be seen as the characteristic function of a set $X \subseteq \mathbb{N}$, where the n th bit of the oracle string is 1 iff $n \in X$. (Ordinary Turing machines are equivalent to oracle Turing machines with the *null oracle* – an oracle which is written as "00000..."; for such machines, the oracle gives no extra information that the Turing machine does not already have.) (Kozen 1997)

Since an oracle can answer questions about membership in a specific set of the natural numbers, an Oracle Turing machine could compute an infinite number of non-recursive functions. It can (trivially) compute the characteristic function of the oracle set, but it could also incorporate its requests to the oracle into more complex algorithms, allowing computation of non-trivial functions. For example, if the oracle set was the *halting set* (the set containing n iff the n th Turing machine halts at some specific input), then the oracle machine would be able to compute many other functions of interest. In fact, it would compute all recursively enumerable functions. A non-recursive oracle is thus a sufficiently powerful resource extending the powers of the ordinary Turing machines.

A procedure using an oracle (possibly, the null oracle) is called *algorithmic relative to X* if it could, at any point of computation, ask and get the *correct* answers in a *finite time* to questions of the form "is n in X ?" We say that a set A is *recursive in X* , or that A is *Turing reducible to X* ($A \leq_T X$), iff there exists an algorithmic relative to X procedure that can compute the characteristic function of the set A .

If a fixed set X is chosen, all (partial) recursive functions of the usual (without (non-null) oracles) recursive function theory can be replaced by corresponding X -recursive (partial) functions in all definitions and theorems. In this way we obtain a

(fully) *relativized theory*. (Various theories which are only *partially* relativized also can be defined and studied.) In the relativized theory of recursive functions it can be shown that the relation \leq_T is reflexive and transitive. If, for a given set A , we require, in addition, that $A \leq_T X$ and $X \leq_T A$, then we say that $A \equiv_T X$, and the equivalence classes of \equiv_T are called *Turing degrees (or levels) of unsolvability* or *T-degrees*. Turing reducibility is commonly taken to be the most fundamental reducibility, though other forms of reducibility are also available (Rogers 1967).

Once we have the notions of relative computability, relative recursiveness, and degrees of unsolvability in arithmetic, the following *hierarchy* of classes of sets of numbers appears, ordered according to their relative degrees of unsolvability:

$$\Sigma_1^0 = \{\text{r.e. sets}\},$$

$$\Delta_1^0 = \{\text{recursive sets}\},$$

$$\Sigma_{n+1}^0 = \{\text{sets r.e. in some } B \in \Sigma_n^0\},$$

$$\Delta_{n+1}^0 = \{\text{sets recursive in some } B \in \Sigma_n^0\},$$

$$\Pi_n^0 = \{\text{complements of sets in } \Sigma_n^0\}.$$

The classes Σ_n^0 , Π_n^0 , and Δ_n^0 constitute what is known as the *arithmetic hierarchy* (or *Kleene hierarchy*). The important feature of the arithmetic hierarchy is that it allows characterization in terms of *first-order quantification* over natural numbers or strings. If we consider second-order quantification – quantification over functions and relations – we get the so called analytic hierarchy consisting of classes Σ_n^1 , Π_n^1 , and Δ_n^1 . The entire

arithmetic hierarchy is strictly contained in Σ_1^1 , the lowest class in the analytic hierarchy. Elements of Δ_1^1 are called *hyperarithmetical sets* (Kozen 1997, Kleene 1943, 1952, Rogers 1967, Shoenfield 1972, Soare 1987).

Modern-day complexity theory has its roots in the theory of recursive functions and effective computability where the Turing-reducibility, completeness and hardness, and the arithmetical hierarchy all have their counterparts (Karp 1972, Cook 1971, and Stockmeyer 1976). Thus, the complexity class of decision problems solvable by an algorithm in class A with an oracle for a problem in class B is written A^B . For example, the class of problems solvable in polynomial time by a deterministic Turing machine with an oracle for a problem in NP is P^{NP} . (This is also the class of problems reducible by polynomial-time Turing reduction to a problem in NP .) Although it is easy to show that $NP \subseteq P^{NP}$, the question of whether NP^{NP} , P^{NP} , NP , and P are equal remains open.

The notation A^B can also mean the class of problems solvable by an algorithm in class A with an oracle for the *language* B . For example, P^{SAT} is the class of problems solvable in polynomial time by a deterministic Turing machine with an oracle for the satisfiability problem.

Oracle machines can also be used for investigating the relationship between complexity classes P and NP , by considering the relationship between P^A and NP^A for an oracle A . In particular, it has been shown that there exist sparse languages (i.e., set of strings such that the number of string with length n in the language is bounded by a polynomial function of n) A and B such that $P^A = NP^A$ and $P^B \neq NP^B$ (Baker *et al.* 1975). The fact that the $P = NP$ question relativizes both ways is taken as evidence that

answering this question will be difficult because any proof technique that *relativizes* (i.e., is unaffected by the addition of an oracle) will not answer the $P = NP$ question.

4.3.2 Oracles and Hypercomputation

Turing, who first introduced the oracles to the recursion theory, left no indication of how and whether these oracles might *actually* be implemented. The only specification he left was that an oracle works by "unspecified means" and "we shall not go any further into the nature of [an] oracle".

The possibility of performing an infinitely many actions, steps, or operations in a finite time has long been a problem of great interest (and confusion) for philosophers since the time of Zeno of Elea. It was James F. Thomson (1954) who coined the term "supertasks" to designate such tasks, and Peter Clarke and Stephen Read (1984) introduced the term "hypertasks" (as well as "super-dupertasks") to designate supertasks with uncountably many steps. Though Thomson himself emphatically denied the mere *logical* possibility of supertasks (as later did Clarke and Read, only with respect to hypertasks), Paul Benacerraf (1962) demonstrated that no cogent argument on purely *logical* grounds had yet shown that a supertask could not be performed. Though most of the discussions in philosophy on supertasks nowadays come from descendants of Benacerraf who accept the logical consistency of the very notion of a supertask, philosophers who reject this possibility tend not to do it on grounds such as Thomson's. Rather, they may appeal to the problems with the notion of infinity itself, or the problems of a particular formal theory within which the corresponding models are formulated. For example, William McLaughlin (1998) claimed that Thompson's

arguments fail to demonstrate the logical possibility of supertasks if analyzed with internal set theory, a variant of real analysis.

Yet, the prevailing view nowadays takes it that, in words of Earman and Norton (1996), our notions of infinity and continuity are now so well developed that supertask have lost their power to force us to *refine* these notions; any difficulties or contradictions that supertasks may deliver no longer reveal deficiencies in our concepts and they can be removed without requiring us to assume some conceptual incoherence in the very notion of supertask.

As a result, a number of various proposals of *conceptual* machines capable of performing supertasks have been proposed: accelerating infinite machines such as Zeus machines, Weyl machines, π -machines, Davies' building infinity machines, (see, e.g., Weyl 1927, Earman and Norton 1993, Davies 2001), machines operating in special relativistic spacetimes (the Pitowsky-Malament-Hogarth spacetimes) harnessing time dilation effects (Pitowsky 1990, Earman and Norton 1993, Hogarth 1992, 1994, Earman 1995), asynchronous neural networks (Copeland and Sylvan 1999, Siegelmann 1998), etc. However, the possibility of *physical* implementation of these machines remains the Achilles' heel of these proposals due to clearly unphysical nature of many essential assumptions used in those models.

Within the theory of computation, a similar discussion resolves around the notion of *hypercomputation*. This term, coined by Jack Copeland (as is, e.g., in 1998, see also Copeland and Proudfoot 1999, Copeland and Sylvan 1999, Ord 2002, 2006), refers to various proposed methods of computing non-Turing computable functions, typically emphasizing the *physical* possibility of implementing such devices.

Recently, Tien D. Kieu, in a number of papers (2002, 2003, 2004, 2005) claimed to have a scheme, according to which, in principle, a real physical *quantum* system could be used to compute prototypical non-Turing computable function in a *finite time*. To our knowledge, this has been the first rigorous attempt to utilize the peculiarities of the quantum world for the project of hypercomputation. In the sections that follow, after exposing the proposed algorithm, I will critically address the proposal showing its *failure* to perform the purported hypercomputation. Whether or not the class of physically realizable hypercomputers is non-empty, Kieu's quantum adiabatic algorithm is not the member of this distinguished club.⁵⁰

4.3.3 The Quantum Adiabatic "Hypercomputer"

4.3.3.1 Introduction

Kieu's insight was to harness the quantum adiabatic algorithm of Farhi *et al.* to solve another decision problem, namely, Hilbert's Tenth. His idea was that one can capitalize on the infinite dimensionality of the Hilbert space that "accompanies" every quantum system in order to perform *in parallel* infinite computational steps in a finite time – a task that a hypercomputer, whether classical or quantum, is supposed to be capable of performing.

Kieu designed the target (interpolated) Hamiltonian as to mimic the form of the left-hand-side squared of the original Diophantine equation. This, in turn, guaranteed the existence of a global minimum: The Diophantine equation has at least one integer solution if the final ground state of the target Hamiltonian is zero, and has no integer

⁵⁰ This arguments contained in this part of the thesis were the results of the joint work that we did in collaboration with Amit Hagar (Hagar and Korolev 2006, 2007a, 2007b).

solutions otherwise. Next, Kieu claimed to have proven an ingenious probabilistic criterion that allows one, by measuring H_p , to identify whether the quantum system has indeed reached its ground state, no matter what T is.⁵¹ If not, according to Kieu, one needs only to enlarge the evolution time T and iterate the algorithm many times, until the ground state (which is ensured to exist through the boundedness of H_p) is achieved.

Let us consider a particular example, say, the following Diophantine equation:

$$D(x, y, z) = (x+1)^3 + (y+1)^3 - (z+1)^3 + cxyz = 0, \quad c \in \mathbb{Z},$$

with unknowns x , y , and z . To find out whether this equation has any non-negative integer solution by a quantum algorithm, it requires the realization of a Fock space. Upon this Hilbert space, we construct the Hamiltonian corresponding to the last expression:

$$H_p = ((a_x^\dagger a_x + 1)^3 + (a_y^\dagger a_y + 1)^3 - (a_z^\dagger a_z + 1)^3 + c(a_x^\dagger a_x)(a_y^\dagger a_y)(a_z^\dagger a_z))^2,$$

which has a spectrum bounded from below – semidefinite, in fact.⁵²

Note that the operators $N_j = a_j^\dagger a_j$ have only non-negative integer eigenvalues n_j , and that $[N_j, H_p] = 0 = [N_j, N_j]$ so these observables are simultaneously measurable.

For some triple (n_x, n_y, n_z) the ground state $|g\rangle$ of the Hamiltonian so constructed has the properties

⁵¹ According to Kieu (e.g., in his 2005, 178), this criterion amounts to excluding any state other than the ground state from occupying the energy spectrum of H_p with probability $> 1/2$ for any $T > 0$. It is noteworthy that in all of his papers Kieu offered no analytic proof for this criterion, only a simple example in which such criterion is indeed satisfied.

⁵² The creation operators a_j^\dagger are similar to those of the 3-D simple harmonic oscillator.

$$\begin{aligned} [a_j, a_j^\dagger] &= 1 \quad \text{for } j = x, y, z, \\ [a_k, a_j^\dagger] &= [a_k, a_j] = 0 \quad \text{for } j \neq k. \end{aligned}$$

$$N_j |g\rangle = n_j |g\rangle,$$

$$H_p |g\rangle = ((n_x + 1)^3 + (n_y + 1)^3 - (n_z + 1)^3 + cn_x n_y n_z)^2 |g\rangle \equiv E_g |g\rangle.$$

Thus, after enough iterations, a projective measurement of the energy E_g of the ground state $|g\rangle$ will yield the answer for the decision problem: the Diophantine equation has at least one integer solution if $E_g = 0$, and has no solutions otherwise. (If $c = 0$ in our example, we know that $E_g > 0$ from Fermat's last theorem.)

If there is one unique solution, then the projective measurements of the observables corresponding to the operators N_j will reveal the values of various unknowns. If there are many solutions, finitely or infinitely as in the case of the Pythagoras theorem, $x^2 + y^2 - z^2 = 0$, the ground state $|g\rangle$ will be a linear superposition of states of the form $|n_x\rangle |n_y\rangle |n_z\rangle$, where (n_x, n_y, n_z) are the solutions. In such a situation, the measurement may not yield all the solutions. However, finding all the solutions is not the aim of a decision procedure for this kind of problem.

Notwithstanding this, measurements of N_j of the ground state would always yield some values (n_x, n_y, n_z) , and a straightforward substitution would confirm whether the equation has a solution or not. Thus the measurement on the ground state either of the energy or of the number operators will be sufficient to give the result for the decision problem.

Since the final Hamiltonian (designed as to mimic the left-hand-side squared of the original Diophantine equation) has an integer spectrum and is bounded from below

(i.e., there exists, by construction, a global minimum for H_p), the evolution time of Kieu's algorithm is finite. Thus, it appears that, at least in theory, Kieu's hypercomputer does indeed work: Given that the algorithm purports to find a global energy minimum, all one needs to do in order to compute the (recursive-theoretic) non-computable is to let the system evolve *slowly enough*, measure its energy, and iterate this procedure until a ground state is achieved with probability $> 1/2$ and an answer to the decision problem is found.

A major breakthrough in computer science? A vindication of the superiority of quantum computers over their classical counterparts? Unfortunately, neither is true. The next section explains why.

4.3.3.2 How Slowly is Slowly Enough?

I now proceed to show that the proposed quantum adiabatic algorithm cannot solve a recursive-theoretic non-computable problem.

A crucial ingredient in the adiabatic algorithm is the energy gap between the ground state E_0 and the next excited state E_1 :

$$g_{\min} = \min_{0 \leq t \leq T} (E_1(t) - E_0(t)).$$

This gap controls the evolution time of the algorithm, in the exact following way:

$$T \gg \mathcal{E} / g_{\min}^2,$$

where

$$\mathcal{E} = \max_{0 \leq s \leq 1} \left| \langle l=0; s | \frac{d\tilde{H}}{dx} | l=0; s \rangle \right|.$$

By making

$$|\langle l=0; s=1 | \psi(T) \rangle|$$

arbitrary close to 1, we obtain that the size of T is governed by the following condition:

$$T \gg 1/g_{\min}^2.$$

The problem is that in the absence of a detailed spectral analysis, *in general* nobody knows what g is, how it behaves, or how to compute it!

Now some of the fanfare in Kieu's papers is built around the idea that there *always* exists such a gap and that the computation *halts* in any case (since the final Hamiltonian H_p , by construction, has an integer spectrum and is bounded from below). We set aside the issue of the feasibility of the manufacturing of such a Hamiltonian, which appears to require infinite precision (Hodges 2005), but even if we grant such (possibly unrealizable) manufacturing capacities, their merit is still questionable: Classically too there may always exist a halting time, only that it is not computable. This is easiest to appreciate in the case of classical Turing's halting problem: Consider all Turing machines with k states; throw away all those that fail to stop on the input 1; among the others take the one that runs longest; call the number of steps of that machine $T(k)$. Now we "know" that in order to decide whether a machine with k states stops on the input 1, we have to wait $T(k)$ steps. But of course we don't really know, because $T(k)$ is not computable, growing faster than any recursive function.

What Kieu is doing is defining an adiabatic process whose time is of that order (and whose gap g is therefore uncomputably small). The fact that there is some T which will do the job is not a big deal (nor is, therefore, the fact that we can use finite but

unbounded dimensional Hilbert spaces for each instance). Indeed, if someone told us what $T(k)$ is, we would not have needed infinitely many steps to complete the job.

With this gap in mind, we can now think of the following problem: For each given running-time of the algorithm T we have to come up with a process whose rate of change is $\sim T^{-1}$. Question: How do we know that we are implementing the correct rate of change *while $H(t)$ is evolving*? Apparently, by being able to measure differences of order T^{-1} , that is, having a sensitive “speedometer”. When the going gets rough we approach very slow speeds of the order of $T^{-1}(k)$, which begs the question, since we can then compute $T(k)$ using our “speedometer”; no fancy quantum computer is needed. If we don't have a “speedometer”, then even if we decided to increase the running-time from T to, say, $T+7$, we will have no clue that the machine is indeed doing $(T+7)^{-1}$ km/h and not T^{-1} km/h. In this case, clearly, Kieu's algorithm cannot be implemented since we will never know how slowly we should evolve the physical system. But then we will also fail to fulfill the adiabatic condition which ensures that once we have reached the desired final Hamiltonian its ground state encodes the solution.

Kieu may argue in response that his (allegedly proven) ingenious probabilistic criterion (along with the iteration of the algorithm) allows him to detect whether the ground state was achieved, that is, whether the algorithm has indeed evolved adiabatically so to ensure a meaningful result when reading off the energy eigenstate of H_p . His idea seems to be the following: *In general*, when one performs such an adiabatic cooling, one doesn't meet this probabilistic criterion (that ensures that it is only the ground state which will appear with probability $> 1/2$ upon measuring H_p) for *any*

state – applying the number operator gives many different answers when one repeats the experiment, and none of them comes up more than half of the time. In this case one simply doubles the running time and tries again, and so on. Since Kieu claimed, call it the HALF CLAIM, that no matter what the value of $T > 0$, no non-ground "decoy" state could ever achieve an occupation probability greater than $1/2$, the algorithm is bound to succeed eventually.

Now if the HALF CLAIM were true, it would have been a remarkable achievement. To see this, recall that the adiabatic theorem provides only a *sufficient* condition for tracking the ground state. In other words, it only guarantees that the system's evolution will track the ground state when certain conditions are met (and only in the adiabatic limit, i.e., when $T \rightarrow \infty$). By claiming that, *no matter what T is*, no state other than the ground state will occupy the energy spectrum with probability $> 1/2$, Kieu is in fact claiming to have proven a theorem which is much stronger than the adiabatic theorem, which all by itself says nothing about non-adiabatic evolutions.

Intuitively, then, it would not be at all surprising if the HALF CLAIM turned out to be false. And unfortunately, as it turns out, the HALF CLAIM *is* false. Although it is true in the adiabatic limit (when $T \rightarrow \infty$) and for a finite T in very special (and very simple) cases of two- and three-dimensional Hamiltonians (which happen to be those picked by Kieu in his numerical simulations that accompanied the HALF CLAIM), it turns out that, for a finite T , some "decoy" excited states may occupy the energy spectrum with much higher probability than the desired ground state of H_p in dimensions higher than three. Indeed, Smith in (2005) constructed several counterexamples to the HALF CLAIM, thus proving its falsity. Interestingly, Smith

claims that one of his counterexamples considers a 5-state system H_I and H_P which exactly (up to a truncation of all matrices down to 5 dimensions) agree with those arising from Kieu's construction (for a certain 1-variable Diophantine problem) and exactly in his " $|n\rangle$ basis" as follows. Consider the following H_I and H_P in a 5-state basis $\{|0\rangle, |1\rangle, |2\rangle, |3\rangle, |4\rangle\}$:

$$H_I = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -\sqrt{2} & 0 & 0 \\ 0 & -\sqrt{2} & 3 & -\sqrt{3} & 0 \\ 0 & 0 & -\sqrt{3} & 4 & -\sqrt{4} \\ 0 & 0 & 0 & -\sqrt{4} & 5 \end{pmatrix}, \quad H_P = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The eigenvectors of H_I are the columns of

$$Q_I = \begin{pmatrix} -0.6198 & -0.6541 & -0.4122 & 0.1336 & 0.0162 \\ -0.6127 & 0.08554 & 0.6350 & -0.4527 & -0.0959 \\ -0.4232 & 0.5151 & 0.0487 & 0.6701 & 0.3226 \\ -0.2300 & 0.4861 & -0.5055 & -0.1675 & -0.6536 \\ -0.0922 & 0.2513 & -0.4111 & -0.5478 & 0.6777 \end{pmatrix}$$

with corresponding eigenvalues (energy values):

$$0.0114, 1.1307, 2.5406, 4.3884, 6.9288.$$

After time evolution, starting from the ground state of H_I at $t = 0$, to $t = T = 13.3444$ we get the following final state $\vec{\psi}$ (absolute values of the $\vec{\psi}$ entries are shown, ordered in the same order as the energies 1, 2, 3, 4, 5 of the H_P eigenstates):

$$(0.0139, 0.9997, 0.0062, 0.0210, 0.0015).$$

Note that this final state has probability $> 99.9\%$ of being measured as the first *excited* state of H_1 , with energy 2, instead of the ground state with energy 1. The expected final energy is 2.007.⁵³

Now, if the HALF CLAIM is false, then the dream of the quantum adiabatic hypercomputer evaporates.

4.3.3.3 The Same Old Story (Told Quantum Mechanically)

In order to see what is left of the quantum adiabatic hypercomputer, stripped as it is from the HALF CLAIM, let us first remind ourselves what undecidability means in the ordinary classical regime.

Suppose we have a black box implementing some function (unknown to us); it takes natural numbers as input and produces natural numbers as output according to some rule hidden inside the box. The designers of the box have assured us that the function is bounded from below, namely, it has a global minimum. Assuming that all we can do is to call this function (use the black box) as many times as we wish (plus some thinking), is it possible to find the function's global minimum? The answer is clearly no, but it is instructive to see exactly why.

In trying to locate a global minimum we can proceed either systematically, by going over each consecutive natural number starting from 0, feeding it into the box and recording the corresponding output, or in some more complicated deterministic or probabilistic manner. At each step, out of all arguments we have checked so far we keep those that minimize the function and discard the others; the former are the global

⁵³ For more details and other counterexamples refer to Smith (2005).

minimum candidates. Note that, if we proceed systematically, sooner or later, after a finite number of steps (number of function's callings), we will always reach a function's global minimum (as we know a global minimum exists). This knowledge (of the fact that we will eventually stumble upon a global minimum), however, adds next to nothing to solving our task. The problem, obviously, is that, even if we have just reached an actual (non-zero) global minimum, there is no way for us to identify it as such. Given the resources we have, we can never be sure whether the function does not take a yet smaller value on the next step.

Thus the fact that we will always reach a global minimum in a finite number of steps is of no help to us. The problem is undecidable only due to our principal inability to identify a global minimum as such. Logically, the reason for this undecidability is that defining the property of being a *global* minimum involves quantification over an infinite domain: we say that the function f reaches its global minimum at a point $n_0 \in N$ iff

$$\forall n \in N : f(n_0) \leq f(n).$$

Trying to identify a global minimum as such by brute-force search would require checking the inequality infinitely many times, hence undecidability.

Coming back to Kieu's proposal, while guaranteeing that the brute-force search will eventually halt, Kieu fails to supply a criterion that would allow one to identify whether or not the algorithm has halted on the global minimum. Consequently, the whole construction, despite the aspirations, lacks the ability to *identify* a global minimum as *global* minimum. The problem is thus no different from the typical classical case of undecidability considered above, and quantum mechanics adds nothing to its solution.

Put another way, the gist behind the adiabatic algorithm is that, after a sufficiently long evolution time, one can be certain to have retrieved the correct result of the decision problem just by performing a measurement on the ground state. However, when the evolution time is unknown, a non-zero energy reading upon a measurement of a final state can be interpreted in two very different ways. On one hand, it may be said to be an eigenvalue of an excited state. In such case, clearly, the evolution was non-adiabatic, hence one must iterate the algorithm with another, longer, evolution time. On the other hand, it may be said to be an eigenvalue of the ground state. In such case, clearly, the algorithm has performed correctly, and one has a (negative) answer to the decision problem. But since one cannot check a negative answer to a classically undecidable problem, how can one tell, without knowing T in advance, that this negative "answer" is indeed correct – that is, that no iterations are needed anymore? Without a criterion for distinguishing a ground state from all other excited states which is *independent* of the knowledge of the adiabatic evolution time T , one simply cannot.

So, Kieu's quantum adiabatic "hypercomputer" fails for a simple reason: As one should intuitively expect from an algorithm that relies on the adiabatic theorem alone, even if the adiabatic conditions are satisfied, then for a *finite* running time $t < T$, there is *in general* no guarantee that the final energy state will be the ground state. Consequently, there is no way to distinguish a "decoy" excited state from a non-zero ground state, i.e., there is no way to identify a global minimum *as such*. Repairing this failure requires knowing in advance the exact adiabatic running time T (or, equivalently, the precise behaviour of the energy gap through out the time-evolution of the algorithm), which is just another undecidable problem.

5. Conclusion

In this thesis I presented three case studies investigating in principle constraints on predictability of the behaviour of physical mechanistic systems in classical and quantum settings. I started with examining the sources of indeterminism and acausality in classical physics that underlie *ontological* constraints on predictability (Part I). Here I discussed the role and physical significance of a Lipschitz condition – a condition violation of which leads to generation of stochastic anomalous motion in the Norton-type indeterministic systems. I argued that the singularity arising from the violation of the Lipschitz condition in the systems considered appears to be so fragile as to be easily destroyed by slightly relaxing certain (infinite) idealizations pertaining to elastic properties of bodies that are required by these models. As a result, I argued that indeterminism of the Norton-type Lipschitz-indeterministic systems should rather be viewed as an artefact of certain (infinite) idealizations essential for the models, depriving the examples of much of their intended metaphysical import, as, for example, in Norton's antifundamentalist programme.

In Part II of the thesis I examined the predictive computational limitations of a classical Laplace's demon. I demonstrated that, in situations that allow self-fulfilling prognoses to take place, the class of undecidable propositions about certain future events, in general, is not empty; any Laplace's demon having *all* the information about the world now will be, in general, *unable* to predict all the future. In order to answer certain questions about the future it needs to resort occasionally to, or to consult with, a demon of a higher order in the computational hierarchy whose computational powers are beyond that of any Turing machine – an *Oracle*. I also discussed the distinction between

ontological and *epistemological* views of determinism, and how adopting Wheeler-Landauer view of physical laws can entangle these aspects on a more fundamental level.

Finally, in Part III, I examined a recent proposal to certain quantum adiabatic algorithm to perform *hypercomputation*. If implemented, a device realizing such an algorithm could serve as a physical realization of an Oracle needed for a Laplacian demon to accomplish its job, and, presumably, seriously damage our traditional views on the limits of predictability and the limits of mathematical (or, more generally, rational knowledge). I critically reviewed this proposal pointing out its failure to deliver the purported hypercomputation. Regardless of whether the class of physically possible hypercomputers is non-empty, Kieu's proposed algorithm is not a member of this distinguished club, and a quantum computer powered Laplace's demon can do no more than its ordinary classical counterpart, retaining the traditional limits of predictability.

BIBLIOGRAPHY

- Abrams, D. S. and Lloyd, S. (1998), "Nonlinear Quantum Mechanics Implies Polynomial-time Solution for NP-complete and #P Problems", Preprint Archive: <http://arxiv.org/quant-ph/9801041>.
- Adleman, L. M. and Blum, M. (1991), "Inductive Inference and Unsolvability", *Journal of Symbolic Logic* 56(3), pp. 891-900.
- Adleman, L. M. and Manders, K. L. (1976), "Diophantine Complexity", in *17th Annual Symposium on Foundations of Computer Science*, Houston, Texas, October 25-26, pp. 81-88.
- _____ (1978), "NP-complete Decision Problems for Binary Quadratics", *Journal of Computer and System Sciences* 16(2), pp. 168-184.
- Adler, I. (1974), *Thinking Machines*, New York: The John Day Company.
- Aharonov, D. (1998), "Quantum Computing", in *Annual Review of Computational Physics VI*, Singapore: World Scientific. See also Preprint Archive: <http://arxiv.org/quant-ph/9812037>.
- Aharonov, Y. and Rohrlich, D. (2005), *Quantum Paradoxes: Quantum Theory for the Perplexed*, Weinheim: Wiley-VCH.
- Aharonov, Y. and Vaidman, L. (2001), "The Two-State Vector Formalism of Quantum Mechanics", Preprint Archive: <http://arxiv.org/quant-ph/0105101>.
- Anderson A. R. (1970), "St. Paul's Epistle to Titus", in Martin (1970), pp. 1-11.

- Angluin, D. and Smith C. H. (1983), "Inductive Inference: Theory and Methods",
Computing Surveys 15(3), pp. 237-269.
- Arnold, V. I. (1978), *Mathematical Methods of Classical Mechanics*, New York,
 Heidelberg, Berlin: Springer-Verlag. Translated from the Russian by K.
 Vogtmann and A. Weinstein.
- _____ (1992), *Ordinary Differential Equations*, Berlin, Heidelberg, New York,
 London, Paris, Tokyo, Hong Kong, Barcelona, Budapest: Springer-Verlag.
 Translated from the Russian by Roger Cooke.
- Arnold, V. I. and Avez, A. (1968), *Ergodic Problems of Classical Mechanics*, New
 York: Benjamin. Translated from the Russian.
- Baker, T. P., Gill, J., Solovay, R. (1975), "Relativizations of the $P = ? NP$ Question",
SIAM Journal on Computing 4(4), pp. 431-442.
- Barrow, J. D. (1990), *The World Within the World*, Oxford, New York: Oxford
 University Press.
- Batterman, R. W. (1993), "Defining Chaos", *Philosophy of Science* 60, pp. 43-66.
- _____ (2002), *The Devil in the Details*, Oxford University Press.
- Batterman, R. W. and White, H. (1996), "Chaos and Algorithmic Complexity",
Foundations of Physics 26, pp. 307-336.
- Benacerraf, P. (1962), "Tasks, Super-Tasks, and the Modern Eleatics", *Journal of*
Philosophy 59, pp. 765-784.
- Bernstein, J. (1991), *Quantum Profiles*, Princeton: Princeton University Press.

- Biham, E., *et al.* (2004), "Quantum Computing Without Entanglement", *Theoretical Computer Science* 320, pp. 15-33.
- Bishop, R. C. (2002), "Deterministic and Indeterministic Descriptions", in Atmanspacher, A. and Bishop, R. C. (eds.), *Between Chance and Choice*, Thorverton: Imprint Academic, pp. 5-31.
- _____ (2003), "On Separating Predictability and Determinism", *Erkenntnis* 58, pp. 169-188.
- _____ (2004), "Anvil or Onion? Determinism as a Layered Concept", *Erkenntnis* 63(1), pp. 55-71.
- Boolos, G. S. and Jeffrey, R. C. (1989), *Computability and Logic*, Cambridge, New York: Cambridge University Press.
- Boyd, R. (1972), "Determinism, Laws and Predictability in Principle", *Philosophy of Science* 39, pp. 431-456.
- Bricmont, J. (1995), "Science of Chaos or Chaos in Science", *Physicalia Magazine* 17, pp. 159-208.
- Bunge, M. (1967), *Foundations of Physics*, Berlin: Springer-Verlag.
- Calude, C., Campbell, D. I., Svozil, K., and Ștefănescu, D. (1994a), "Strong Determinism vs. Computability", Preprint Archive: <http://arxiv.org/quant-ph/9412004>.
- Calude, C., Jürgensen, H. and Zimand, M. (1994b), "Is Independence an Exception?", *Appl. Math. Comput.* 66, pp. 63-76.

Campbell, J. K., O'Rourke, M., and Shier, D. (eds.) (2004), *Freedom and Determinism*, Cambridge, Massachusetts: MIT Press.

Cantor, G. (1874), "Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen", *J. für die reine und angewandte Mathematik* 77, pp. 258-262. Reprinted in *George Cantor Gesammelte Abhandlungen*, Berlin: Springer-Verlag, 1932, pp. 115-118.

Chaitin, G. J. (1987), *Algorithmic Information Theory*, Cambridge: Cambridge University Press.

Church, A. (1933), "A Set of Postulates for the Foundation of Logic", *Ann. Math.* 33-34, pp. 346-366, 839-864.

_____ (1936), "An Unsolvable Problem of Elementary Number Theory", *Amer. J. Math.* 58, pp. 345-363.

Clarke, P. and Read, S. (1984), "Hypertasks", *Synthese* 61(3), pp. 387-390.

Conway, J. H. and Guy, R. K. (1996), *The Book of Numbers*, New York: Springer-Verlag.

Cook, S. A. (1971), "The Complexity of Theorem Proving Procedures," in *Proc. 3rd Ann. ACM Symp. on Theory of Computing*, Association for Computing Machinery, pp. 151-158.

Copeland, J. (1998), "Super Turing Machines", *Complexity* 4, pp. 30-32.

_____ (2002), "Hypercomputation", *Minds and Machines* 12, pp. 281-301.

- _____ (2004), *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, Plus The Secrets of Enigma*, Oxford, UK: Clarendon Press; New York: Oxford University Press.
- Copeland, J. and Proudfoot, D. (1999), "Alan Turing's Forgotten Ideas in Computer Science", *Scientific American* 280, pp. 99-103.
- Copeland, J. and Sylvan, R. (1999), "Beyond the Universal Turing Machine", *Australasian Journal of Philosophy* 77, pp. 46-66.
- Curry, H. B. (1929), "An Analysis of Logical Substitution", *Amer. J. Math.* 51, pp. 363-384.
- Cushing, J. (1990), *Theory Construction and Selection in Modern Physics: The S-Matrix*, Cambridge: Cambridge University Press.
- Da Costa, N. C. A. and Doria, F. A. (1991a), "Classical Physics and Penrose's Thesis", *Foundations of Physics Letters* 4(4), pp. 363-373.
- _____ (1991b), "Undecidability and Incompleteness in Classical Mechanics", *International Journal of Theoretical Physics* 30, pp. 1041-1073.
- Davis, M. (1953), "Arithmetical Problems and Recursively Enumerable Predicates", *Journal of Symbolic Logic* 18(1), pp. 33-41.
- _____ (1958), *Computability and Unsolvability*, New York: McGraw-Hill.
- _____ (1965), *The Undecidable*, New York: Raven Press.
- _____ (1973), "Hilbert's Tenth Problem is Unsolvable", *Amer. Math. Monthly* 80, pp. 233-269.

- _____ (2003), "The Myth of Hypercomputation", in Teuscher, C. (ed.) *Alan Turing, Life and Legacy of a Great Thinker*, New York: Springer.
- Davies, E. B. (2001), "Building Infinite Machines", *British Journal for the Philosophy of Science* 52, pp. 671-682.
- Davies, P. C. W. (ed.) (1989), *The New Physics*, Cambridge, England: Cambridge University Press.
- _____ (1994), "Emergent Biological Principles and the Computational Properties of the Universe". Preprint Archive: <http://arxiv.org/astro-ph/0408014>.
- Demopoulos, W. (2003) "Elementary Propositions and Essentially Incomplete Knowledge: A Framework for the Interpretation of Quantum Mechanics", *Noûs* 38(1), pp. 86-109.
- Denef, J., et al. (eds.) (2000), *Hilbert's Tenth Problem: Relations with Arithmetic and Algebraic Geometry, Workshop on Hilbert's Tenth Problem*, November 2-5, 1999, Ghent University, Belgium, American Mathematical Society: Providence, Rhode Island.
- Diacu, F. and Holmes, P. (1996), *Celestial Encounters: The Origins of Chaos and Stability*, Princeton, N.J.: Princeton University Press.
- Earman, J. (1986), *A Primer on Determinism*, Dordrecht, Boston, Lancaster, Tokyo: D. Reidel Publishing Company.
- Earman, J. (1995), *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes*, New York, Oxford: Oxford University Press.

- _____ (2004), "Determinism: What We Have Learned and What We Still Don't Know", in Campbell *et al.* (2004), pp. 21-46.
- Earman, J. and Norton, J. (1993), "Forever is a Day: Supertasks in Pitowsky and Malament-Hogarth Spacetimes", *Philosophy of Science* 60(1), pp. 22-42.
- _____ (1996), "Infinite Pains: The Trouble with Supertasks", in *Benacerraf and His Critics*, Morton, A. and Stich, S. P. (eds.), Cambridge, Massachusetts: Blackwell Publishers.
- Enderton, H. (1977), "Elements of Recursion Theory", in *Handbook of Mathematical Logic*, Barwise, J. (ed.), Amsterdam: North-Holland Publishing Co., pp. 527-566.
- Epstein, R. L. and Carnielli, W. A. (2000), *Computability: Computable Functions, Logic, and the Foundations of Mathematics*, Wadsworth: Australia, etc., 2nd ed.
- Eslami, M. (1994), *Theory of Sensitivity of Dynamical Systems: An Introduction*, Berlin, New York: Springer-Verlag.
- Farhi, E., *et al.* (2000a), "Quantum Computation by Adiabatic Evolution", Preprint
Archive: <http://arxiv.org/quant-ph/0001106>.
- Farhi, E., *et al.* (2000b), "A Quantum Adiabatic Evolution Algorithm Applied to Random Instances of an NP-Complete Problem", *Science* 20, Vol. 292. No. 5516, pp. 472- 475.
- Feferman, S. (1984), "Kurt Gödel: Conviction and Caution", *Philosophia Naturalis* 21, pp. 546-562.
- Feigl, H. (1953), "Notes on Causality", in Feigl, H. and Brodbeck, M. (1953).

- Feigl, H. and Brodbeck, M. (eds.) (1953), *Readings in the Philosophy of Science*, New York: Appleton-Century-Crofts.
- Feinberg, G., Lavine, S., and Albert, D. (1992), "Knowledge of the Past and Future", *Journal of Philosophy* 89, pp. 607-642.
- Feynman, R. (1982), "Simulating Physics with Computers", *International Journal of Theoretical Physics*, 21, pp. 467-488.
- Ford, J. (1981), "How Random is a Coin Toss?", in Horton, C.W., Reichl, L. E., and Szebehely, V. G. (eds.), *Proceedings of the Workshop on Long-Time Predictions in Non-Linear Conservative Dynamical Systems*, New York: Wiley, pp. 79-92.
- _____ (1983), "How Random is a Coin Toss?", *Physics Today* 36, pp. 40-47.
- _____ (1988), "Quantum Chaos, Is There Any?", in Bai-Lin, H. (ed.), *Directions in Chaos 2*, Singapore: World Scientific, pp. 128-147.
- _____ (1989), "What is Chaos, That We Should be Mindful of It?", in Davies (1989), pp. 348-371.
- Fortnow, L. (2003), "One Complexity Theorist's View of Quantum Computing", *Theoretical Computer Science* 292, pp. 597-610.
- Frigg, R. (2005), "Entropy – One World, Many Concepts", in *Lecture Materials for the Philosophy, Probability, and Physics Summer School 2005*, University of Konstanz, 2005.
- Garey, M. R. and Johnson, D. S. (1979), *Computers and Intractability: A Guide to the Theory of NP-Completeness*, New York: W. H. Freeman & Co.

- Gentzen, G. (1936), "Die Widerspruchfreiheit der reinen Zahlentheorie", *Mathematische Annalen* 112, pp. 493-565. Translated as "The Consistency of Arithmetic", in Szabo M. E. (ed.) (1969), *The Collected Works of Gerhard Gentzen*, Amsterdam: North-Holland.
- Gilmore, R. (1981), *Catastrophe Theory for Scientists and Engineers*, New York, Chichester, Brisbane, Toronto: John Wiley & Sons.
- Glymour, C. (1971), "Determinism, Ignorance, and Quantum Mechanics", *Journal of Philosophy* 68, pp. 744-751.
- Gold, E. M. (1967), "Language Identification in the Limit", *Information and Control* 10, pp. 447-474.
- Gödel, K. (1931), *Monatshefte für Mathematik und Physik* 38, 173; English translation in Gödel (1986) and Davis (1965).
- _____ (1986), *Collected Works, Vol. I, Publications 1929-1936*, ed. by S. Feferman, J. W. Dawson, Jr., St. C. Kleene, G. H. Moore, R. M. Solovay, J. van Heijenoort, Oxford: Oxford University Press.
- Grover, L. (1996), "A Fast Quantum Mechanical Algorithm for Database Search", *Proceedings, STOC*, Philadelphia PA, pp. 212-219.
- Hadamard, J. (1898), "Les surfaces à courbures opposées et leurs lignes géodésiques". *J. Math. Pures et Appl.* 4, pp. 27-73.
- Hagar, A. and Korolev, A. (2006), "Quantum Hypercomputation", *Minds and Machines* 16(1), pp. 87-93.

- _____ (2007a), "What is Quantum in Quantum Computing? – Lessons from Two Halting Problems", *Handbook of Quantum Logic, Quantum Structures, and Quantum Computing* (eds. Dov Gabbay, Daniel Lehmann, and Kurt Engesser), Elsevier, forthcoming.
- _____ (2007b), "Quantum Hypercomputation – Hype or Computation?", forthcoming in *Philosophy of Science*.
- Harel, D. (1987), *Algorithms, The Spirit of Computing*, Wokingham, England: Addison-Wesley.
- Haroche, S. and Raimond, J. M. (1996), "Quantum Computing: Dream or Nightmare?", *Physics Today* 8, pp. 51-52.
- Hartmann, S. (2005a), "The World as a Process: Simulations in the Natural and Social Sciences", Pittsburgh PhilSci Archive: <http://philsci-archive.pitt.edu/archive/00002412>.
- _____ (2005b), "Models and Stories in Hadron Physics", Pittsburgh PhilSci Archive: <http://philsci-archive.pitt.edu/archive/00002433/>.
- Heath, T. L. (1964), *Diophantus of Alexandria: A Study in the History of Greek Algebra*, New York: Dover Publications, Inc., 2nd ed.
- Herbrand, J. (1931), "Sur la non-contradiction de l'arithmétique", *Journal für die reine und angewandte Mathematik* 166, pp. 1-8. Translated as "The Consistency of Arithmetic" in van Heijenoort (1967), pp. 620-628.

Hilbert, D. (1899), "Grundlagen der Geometrie", in Teubner, B. G. (ed.), *Festschrift zur Freier der Enthüllung des Gauss-Weber Denkmals in Göttingen*, Leipzig, pp. 3-92.

_____ (1900), "Mathematische Probleme, Vortrag, gehalten auf dem internationalen Mathematiker Kongress zu Paris 1900", *Nachr. K. Ges. Wiss., Göttingen, Math.-Phys. Kl.*, pp. 253-297. English translation: *Bull. Amer. Math. Soc.* 8 (1901-1902), pp. 437-479. Reprinted in *Mathematical Developments arising from Hilbert problems*, Proceedings of Symposia in Pure Mathematics, vol. 28, American Mathematical Society, Browder Ed. (1976), pp. 1-34. See also *Arch. Math. Phys.* 3rd Ser., Vol. 1 (1901), pp. 44-63 and 213-237. See also Hilbert (1935), p. 310.

_____ (1935), *Gesammelte Abhandlungen*, Berlin: Springer, vol. 3, p. 310. Reprinted: New York: Chelsea (1965).

Hintikka, J. (1996), *The Principles of Mathematics Revisited*, Cambridge, New York: Cambridge University Press.

Hogarth, M. (1992), "Does General Relativity Allow an Observer to View an Eternity in a Finite Time?", *Foundations of Physics Letters* 5, pp. 173-181.

_____ (1994), "Non-Turing Computers and Non-Turing Computability", *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 1: Contributed Papers, pp. 126-138.

Hodges, A. (1992), *Alan Turing: An Enigma*, London, U.K.: Vintage, Random House.

_____ (2005), "Can Quantum Computing Solve Classically Unsolvable Problems?", Preprint Archive: <http://arxiv.org/quant-ph/0512248>.

- Hoefer, C. (2004), "Causality and Determinism: Tension, or Outright Conflict?",
forthcoming in *Revista de Filosofia*. Pittsburgh PhilSci Archive: <http://philsci-archive.pitt.edu/archive/00002071/>.
- Hogarth, M. (1992), "Does General Relativity Allow an Observer to View an Eternity in a Finite Time?", *Foundations of Physics Letters* 5, pp. 173-181.
- _____ (1994), "Non-Turing Computers and Non-Turing Computability", *PSA* 94(1), pp. 126-138.
- _____ (2004), "Deciding Arithmetic Using SAD Computers", *British Journal for the Philosophy of Science* 55, pp. 681-691.
- Hunt, G. (1987), "Determinism, Predictability and Chaos", *Analysis* 47, pp. 129-133.
- Jones, J. P. and Matijasevic, Y. V. (1984), "Register Machine Proof of the Theorem on Exponential Diophantine Representation of Enumerable Sets", *Journal of Symbolic Logic* 49(3), pp. 818-829.
- Jozsa, R. (1997), "Entanglement and Quantum Computation", in S. Hugget *et al.* (eds.) *Geometric Issues in the Foundations of Science*, Oxford: Oxford University Press.
See also Preprint Archive: <http://arxiv.org/quant-ph/9707034>.
- Kac, M. (1975), "Some Reflections of a Mathematician on the Nature and the Role of Statistics", *Advances in Applied Probability* 7, Supplement: Proceedings of the Conference on Directions for Mathematical Statistics, pp. 5-11.
- Kanter, I. (1990), "Undecidability Principle and the Uncertainty Principle Even for Classical Systems", *Phys. Rev. Lett.* 64(4), pp. 332-335.

- Karp, R. M. (1972), "Reducibility Among Combinatorial Problems", in *Complexity of Computer Computations*, Miller, R. E. and Thatcher, J. W. (eds.), New York: Plenum Press, pp. 85-103.
- Kellert, S. H. (1993), *In the Wake of Chaos*, Chicago: University of Chicago Press.
- Kieu, T. D. (2002), "Quantum Hypercomputability", *Minds and Machines* 12, pp. 541-561.
- _____ (2003), "Computing the Noncomputable", *Contemporary Physics* 44, pp. 51-71.
- _____ (2004), "A Reformulation of Hilbert's Tenth Problem through Quantum Mechanics", *Proceedings of the Royal Society A* 460, p. 1535.
- _____ (2005), "An anatomy of a Quantum Adiabatic Algorithm that Transcends the Turing Computability", *International Journal of Quantum Information* 3(1), pp. 177-183.
- Kitaev, A. Yu., Shen, A. H., and Vyal'yi, M. N. (2002), *Classical and Quantum Computation*, Graduate Studies in Mathematics, Vol. 47, Providence, Rhode Island: American Mathematical Society.
- Kleene, S. C. (1935), "A Theory of Positive Integers in Formal Logic", *Amer. J. Math.* 57, pp. 153-173, 219-244.
- _____ (1943), "Recursive Predicates and Quantifiers", *Trans. Amer. Math. Soc.* 53, pp. 41-74.
- _____ (1952), *Introduction to Metamathematics*, Princeton, New Jersey: D. van Nostrand Company, Inc.

- _____ (1967), *Mathematical Logic*, New York: Wiley.
- Kleiner, I. (1998), "From Numbers to Rings: The Early History of Ring Theory", *Elem. Math.* 53, pp. 18-35.
- Kolmogorov, A.N. (1941), "Dissipation of Energy in Locally Isotropic Turbulence", *Dokl. Akad. Nauk, SSSR* 31, pp. 538-540.
- Komar, A. (1964), "Undecidability of Macroscopically Distinguishable States in Quantum Field Theory", *Phys. Rev.* 133(B), pp. 542-544.
- Korolev, A. (1998), "On the Truth-Value of Future Contingents", *Models of Cognitive Processes, Computational Systems* 164, Novosibirsk: Sobolev Institute of Mathematics, pp. 62-68 (in Russian).
- _____ (2007), "Indeterminism, Asymptotic Reasoning, and Time Irreversibility in Classical Physics", forthcoming in *Philosophy of Science*. Pittsburgh PhilSci Archive: <http://philsci-archive.pitt.edu/archive/00003003/>.
- Kozen, D. C. (1997), *Automata and Computability*, New York, Berlin, Heidelberg: Springer-Verlag New York Inc.
- Kreisel, G. (1974), "A Notion of Mechanistic Theory", *Synthese* 29, pp. 11-26.
- Lagrange, J. L. (1772), "Démonstration d'un théorème d'arithmétique", *Nouveaux Mémoires de l'Académie royale des Sciences et Belles-Lettres de Berlin, année 1770*, pp. 123-133.
- Laing, R. (1977), "Automaton Models of Reproduction by Self-Inspection", *Journal of Theoretical Biology* 66(3), pp. 437-456.

Landau, L. D. and Lifshitz, E. M. (1986), *Theory of Elasticity*, 3rd ed., Vol. 7 of *Course of Theoretical Physics*, Translated from the Russian by J.B. Sykes and W.H. Reid, Oxford, New York, Beijing, Frankfurt, São Paulo, Sydney, Tokyo, Toronto: Pergamon Press.

_____ (1976), *Mechanics*, 3rd ed., Vol. 1 of *Course of Theoretical Physics*, Translated from the Russian by J.B. Sykes and J.S. Bell, Oxford, New York, Toronto, Sydney, Paris, Frankfurt: Pergamon Press.

Landauer, R. (1967), "Wanted: A Physically Possible Theory of Physics", *IEEE Spectrum* 4, pp. 105-109.

_____ (1986), "Computation and Physics: Wheeler's Meaning Circuit?", *Foundations of Physics* 16(6), pp. 551-564.

Laplace, P. S. (1820), *Théorie analytique des probabilités*, Paris: V. Courcier. Translated by Truscott F. W. and Emory F. L., *A Philosophical Essay on Probabilities*, New York: Dover Publications, 1814/1951.

Laymon, R. (1985), "Idealization and Testing of Theories by Experimentation", in Achinstein, P. and Hannaway, O. (eds.), *Observation, Experiment, and Hypothesis in Modern Physical Science*, Boston, Massachusetts: MIT Press.

_____ (1995), "Idealizations, Externalities, and the Economic Analysis of Law", in Pitt, J. (ed.), *New Directions in the Philosophy of Technology*, Dordrecht: Kluwer Academic Publishers, pp. 185-206.

Levin, L. A. (2003), "The Tale of One-Way Functions", *Problems of Information Transmission (Problemy Peredachi Informatsii)* 39(1), pp. 92-103.

- Li, M., and Vitányi, P.M.B. (1992), "Inductive Reasoning and Kolmogorov Complexity", *Journal of Computer and System Science* 44, pp. 343-384.
- Linden, N. and Popescu, S. (1999), "Good Dynamics Versus Bad Kinematics: Is Entanglement Needed for Quantum Computation?", *Phys. Rev. Lett.* 87(4), 047901. See also Preprint Archive: <http://arxiv.org/quant-ph/9906008>.
- Lloyd, S. (1995), "Quantum Mechanical Computers", *Scientific American* 95(10), pp. 44-50.
- MacKay, D. M. (1960), "On the Logical Indeterminacy of a Free Choice", *Mind* 69(273), pp. 31-40.
- Mackey, G. W. (1974), "Ergodic Theory and Its Significance for Statistical Mechanics and Probability Theory", *Adv. in Math.* 12, pp. 178-268.
- Maltsev, A. I. (1970), *Algorithms and Recursive Functions*, Groningen: Wolters-Noordhof.
- Mañé, R. (1987), *Ergodic Theory and Differentiable Dynamics*, Berlin, New York: Springer-Verlag. Translated from the Portuguese by Silvio Levy.
- Martin, R. L. (ed.) (1970), *The Paradox of the Liar*, New Haven and London: Yale University Press.
- Matiyasevich, Yu. V. (1970), "Enumerable Sets are Diophantine", *Doklady Akademii Nauk SSSR*, 191. English translation *Soviet Math. Dokl.* 11 (1970), pp. 354-358.
- _____ (1993), *Desyataya problema Gilberta*, Moscow: Fizmatlit. English translation: *Hilbert's Tenth Problem*, The MIT Press: Cambridge, Massachusetts, London, England (1993).

- McLaughlin, W. I. (1998), "Thomson's Lamp is Dysfunctional", *Synthese* 116(3), pp. 281-301.
- McMullin, E. (1985), "Galilean Idealization", *Studies in History and Philosophy of Science* 16, pp. 247-273.
- Messiah, A. (1961), *Quantum Mechanics*, Vol. II, New York: Interscience Publishers.
- Moore, C. (1990), "Unpredictability and Undecidability in Dynamical Systems", *Phys. Rev. Lett.* 64, pp. 2354-2357.
- _____ (1991), "Generalized Shifts: Unpredictability and Undecidability in Dynamical Systems", *Nonlinearity* 4, pp. 199-230.
- Moulines, C. U. (1996), "Structuralist Models, Idealization, and Approximation", in Hegselmann *et al.* (eds.), *Modeling and Simulation in the Social Sciences from the Philosophy of Science Point of View (Theory and Decision Theory)*, Dordrecht: Kluwer Academic Publishers, pp. 157-168.
- Nagel, E. (1961), *The Structure of Science: Problems in the Logic of Scientific Explanation*, New York: Harcourt, Brace, and World.
- Nielsen, M. A. and Chuang I. L. (2000), *Quantum Computation and Quantum Information*, Cambridge: Cambridge University Press.
- Norton, J. (2003), "Causation as Folk Science", *Philosopher's Imprint*, Vol. 3, No. 4, <http://www.philosophersimprint.org/003004>, to be reprinted in Price, H. and Corry, R. (eds.), *Causation and the Constitution of Reality*, Oxford University Press.

- Ord, T. (2002), "*Hypercomputation: Computing More Than the Turing Machine* ",
Preprint Archive: <http://arxiv.org/math.LO/0209332>.
- _____ (2006), "The Many Forms of Hypercomputation", *Applied Mathematics and Computation* 178, pp. 143-153.
- Papadimitriou, C. H. (1994), *Computational Complexity*, Reading, MA: Addison-Wesley.
- Parker, Matthew W. (2006), "Computing the Uncomputable, or, The Discrete Charm of Second-Order Simulacra", *Proceedings (Models and Simulations, London, 2006)*: Paris. Pittsburgh PhilSci Archive: <http://philsci-archive.pitt.edu/archive/00002756/>.
- Peres, A. and Zurek, W. H. (1982), "Is Quantum Theory Universally Valid?", *Am. J. Phys.* 50, pp. 807-810.
- Pitowsky, I. (1982), "Substitution and Truth in Quantum Logic", *Philosophy of Science* 49(3), pp. 380-401.
- _____ (1989), *Quantum Probability – Quantum Logic*, New York, London, Paris, Tokyo: Springer-Verlag.
- _____ (1990), "The Physical Church Thesis and Physical Computational Complexity", *Iyyun* 39, pp. 81-99.
- _____ (1996), "Laplace's Demon Consults an Oracle: The Computational Complexity of Prediction", *Studies in the History and Philosophy of Modern Physics* 27(2), pp. 161-180.

_____ (2002), "Quantum Speed-up of Computations", *Philosophy of Science* 69, pp. S168-S177.

Poincaré, H. (1899), *Les méthodes nouvelles de la mécanique céleste*. Tome I. Paris 1892; Tome III. Paris 1899.

Popper, K. (1950), "Indeterminism in Quantum Physics and in Classical Physics", *British Journal for the Philosophy of Science* 1, pp. 117-133.

_____ (1961), *The Poverty of Historicism*, New York, Evanston: Harper Torchbooks, The Academy Library, Harper & Row, Publishers.

_____ (1982), *The Open Universe*, Totowa: Rowman & Littlewood.

Post, E. (1936), "Finite Combinatory Process-Formulation, I", *Journal of Symbolic Logic* 1, pp. 103-105.

_____ (1943), "Formal Reductions of the General Combinatorial Decision Problem", *Amer. J. Math.* 65, pp. 197-215.

Pour-El, M. and Richards, I. (1981), "The Wave Equation with Computable Initial Data such that Its Unique Solution is not Computable", *Advances in Mathematics* 39, pp. 215-239.

Presburger, M. (1930), "Über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt", *Sprawozdanie z I Kongresu matematyków krajów słowiańskich, Warszawa 1929*, pp. 92-101, and supplement, p. 395. Translated as "On the Completeness of a Certain System of Arithmetic of Whole Numbers in which Addition Occurs as the Only Operation", *History and Philosophy of Logic* 12, 1931, pp. 225-233.

- Preskill, J. (1998), "Quantum Computing: Pro and Con", *Proceedings of the Royal Society of London A* 454, pp. 469-486.
- Putnam, H. (1960), "An Unsolvable Problem in Number Theory", *Journal of Symbolic Logic* 25(3), pp. 220-232.
- Radó, T. (1962), "On Non-Computable Functions", *Bell Systems Technical Journal* 41(3), pp. 877-884.
- Redhead, M. (1980), "Models in Physics", *British Journal for the Philosophy of Science* 31: 145-163.
- Reichardt, B. W. (2004), "The Quantum Adiabatic Optimization Algorithm and Local Minima", *Proceedings of the 36th Symposium on Theory of Computing (STOC)*, pp. 502-510.
- Rice, H. G. (1953), "Classes of Recursively Enumerable Sets and Their Decision Problems", *Trans. Amer. Math. Soc.* 89, pp. 25-59.
- _____ (1956), "On Completely Recursively Enumerable Classes and Their Key Arrays", *Journal of Symbolic Logic* 21, pp. 304-341.
- Rogers, H. (1967), *Theory of Recursive Functions and Effective Computability*, New York: MacGraw-Hill. Reprinted in 1987, Cambridge, Massachusetts: MIT Press.
- Russell, B. (1953), "On the Notion of Cause and with Applications to the Free-Will Problem", in Feigl, H. and Brodbeck, M. (eds.) (1953).
- Sagan, C. and Newman, W. (1985), "The Solipsist Approach to Extraterrestrial Intelligence", in Edward Regis, Jr. (ed.), *Extraterrestrials: Science and Alien Intelligence*, Cambridge: Cambridge University Press, pp. 151-161.

- Schmidt, J. H. (1998), "Predicting the Motion of Particles in Newtonian Mechanics and Special Relativity", *Studies in History and Philosophy of Modern Physics* 29, pp. 81-122.
- Schönfinkel, M. (1924), "Über die Bausteine der mathematischen Logik", *Math. Annalen* 92, pp. 305-316.
- Schurz, G. (1996), "Kinds of Unpredictability in Deterministic Systems", in Weingartner, P. and Schurz, G. (eds.) (1996).
- Schuster, H. G. (1984), *Deterministic Chaos: An Introduction*, Weinheim, Federal Republic of Germany: Physik-Verlag; Deerfield Beach, FL, USA: VCH Publishers [distributor].
- Shannon, C. E. and McCarthy, J. (eds.) (1956), *Automata Studies*, Princeton: Princeton University Press.
- Shoenfield, J. R. (1967), *Mathematical Logic*, Reading, Massachusetts: Addison-Wesley.
- _____ (1972), *Degrees of Unsolvability*, New York: Elsevier.
- Shor, P. (1994), "Algorithms for Quantum Computation: Discrete Log and Factoring", *Proceedings of the 35th IEEE Symposium on Foundations of Computer Science*, pp. 124-134.
- Siegelmann, H. (1998), *Neural Networks and Analog Computation: Beyond the Turing Limit*, Boston, Massachusetts: Birkhauser.

- Skolem, T. (1931), "Über einige Satzfunktionen in der Arithmetik", *Skrifter utgitt av Det Norske Videnskaps-Akademi I Oslo, I. Mat.-natur. kl.*, no. 7, pp. 1-28. Reprinted in *Skolem Selected Works in Logic*, Oslo: Universitetsforlaget, 1970, pp. 281-306.
- Sipser, M. (2006), *Introduction to the Theory of Computation*, Boston: Thomson Course Technology.
- Smith, W. (2005), "Three Counterexamples Refuting Kieu's Plan for 'Quantum Adiabatic Hypercomputation'; and Some Uncomputable Quantum Mechanical Tasks", <http://math.temple.edu/~wds/homepage/works.html>, #87, 2005. To appear in *Journal of the Association for Computing Machinery*.
- Smoryński, C. (1991), *Logical Number Theory I: An Introduction*, Berlin: Springer-Verlag.
- Smullyan, R. M. (1994), *Diagonalization and Self-Reference*, Oxford, New York: Clarendon Press.
- Soare, R. I. (1987), *Recursively Enumerable Sets and Degrees*, Berlin: Springer-Verlag.
- Sobel, J. H. (1998), *Puzzles of the Will: Fatalism, Newcomb and Samarra, Determinism and Omniscience*, Toronto: Toronto University Press.
- Sokolnikoff, I. S. (1956), *Mathematical Theory of Elasticity*, 2nd ed., New York, Toronto, London: McGraw-Hill Book Company.
- Steiner, F. (1994), "Quantum Chaos", invited contribution to the *Festschrift Universität Hamburg 1994: Schlaglichter der Forschung zum 75. Jahrsdag* (Ed. R. Ausorge), published on the occasion of the 75th anniversary of the University of Hamburg,

Dietrich Reimer Verlag: Hamburg, pp. 542-564. Preprint Archive:

<http://arxiv.org/abs/chao-dyn/9402001>.

Stockmeyer, L. J. (1976), "The Polynomial-Time Hierarchy", *Theoretical Computer Science* 3, pp. 1-22.

Stone, M. A. (1989), "Chaos, Prediction, and Laplacian Determinism", *American Philosophical Quarterly* 26, pp. 123-131.

Stroock, D. W. and Varadhan, S. R. S. (1969), "Diffusion Processes with Continuous Coefficients", I and II, *Commun. Pure Appl. Math.* 22, pp. 345-400.

_____ (1979), *Multidimensional Diffusion Processes*, Berlin: Springer.

Svozil, K. (1986), "Computational Universes", *Il Nuovo Cimento* 96B, pp. 127-139.

_____ (1993), *Randomness & Undecidability in Physics*, Singapore, New Jersey, London, Hong Kong: World Scientific.

_____ (1995), "Consistent Use of Paradoxes in Deriving Constraints on the Dynamics of Physical Systems and of No-go Theorems ", *Foundations of Physics Letters* 8(6), pp. 523-535.

Tarski, A. (1932), "Der Wahrheitsbegriff in den Sprachen der deduktiven Disziplinen", in *Akademie der Wissenschaften in Wien, Mathematisch-naturwissenschaftliche Klasse, Akademischer Anzeiger* 69, pp. 23-25.

Thomas, T. Y. (1961), *Plastic Flow and Fracture in Solids*, New York, London: Academic Press Inc.

Thomson, J. (1954), "Tasks and Super-Tasks", *Analysis* 15, pp. 1-13.

- Tipler, F. (1981), "Extraterrestrial Beings Do Not Exist", *Quarterly Journal of the Royal Astronomical Society* 21(267).
- Turing, A. M. (1936), "On Computable Numbers, with an Application to the Entscheidungsproble", *Proc. London Math. Soc.* (2) 442, pp. 230-265. Reprinted in Gandy R. O. and Yates C. E. M., (eds.), *Collected Works: Mathematical Logic*, Amsterdam, New York, Oxford, Tokyo: North-Holland, 2001, pp. 18-53.
- _____ (1939), "Systems of Logic Based on Ordinals", *Proc. London Math. Soc.* 42, pp. 230-265.
- Unruh, W. G. (1995), "Maintaining Coherence in Quantum Computers", *Phys. Rev. A* 51, pp. 992-997.
- Van Haijenoort, J. (1967), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, Massachusetts: Harvard University Press.
- Van Kampen, N. G. (1991), "Determinism and Predictability", *Synthese* 89, pp. 273-281.
- Van Leeuwen, J., (ed.) (1990), *Algorithms and Complexity, Volume A*, Cambridge, Massachusetts: Elsevier, Amsterdam and MIT Press.
- Von Neumann, J. (1966), *Theory of Self-Reproducing Automata*, (ed.) by Burks, A. W., Urbana: University of Illinois Press.
- Weinan, E. and Vanden-Eijnden, E. (2003), "A Note on Generalized Flows", *Physica D* 183, pp. 159-174.
- Weingartner, P. and Schurz, G. (eds.) (1996), *Law and Prediction in the Light of Chaos Research*, Berlin: Springer-Verlag.

Weyl, H. (1927), *Philosophie der Mathematik und Naturwissenschaft*, Munich: R. Oldenburg.

Wheeler, J. A. (1983), "On Recognizing 'Law Without Law,'" Oersted Medal Response at the Joint APS-AAPR Meeting, New York, 25 January 1983, *American Journal of Physics* 51(5), pp. 398-404.

Wheeler, J. A. (1986), "Physics as Meaning Circuit: Three Problems", in *Frontiers of Non-equilibrium Statistical Physics*, Moore, G. T. and Scully, M. O., (eds.), New York: Plenum Press.