PREDICTABILITY, COMPUTABILITY, AND SPACETIME

MARK HOGARTH
Sidney Sussex College, Cambridge

A thesis submitted to the Faculty of Philosophy in the University of Cambridge in partial fulfilment of the requirements for the degree of Doctor of Philosophy. The thesis is the result of the author's own work and includes nothing which is the outcome of work done in collaboration.

To my Mum and Dad,
who succeeded in violating Larkin’s Law.
And to my sister Lyn,
who recently stopped pulling my hair.

Acknowledgements

The following have personally helped to shape the ideas in the thesis: Gordon Belot, George Boolos, Tim Budden, Rob Clifton, John Earman, Bob Geroch, David Malament, Joseph Melia, John Norton, Michael Redhead, Adrian Stanley, and Maria Stuart. I thank them all. Special thanks are to due Jeremy Butterfield who supervised this work.

CONTENTS

Preface

Experiment escorts us last —
His pungent company
Will not allow an Axiom
An Opportunity

EMILY DICKINSON

Scholars have differed about the correct interpretation of this poem, though there is evidence that it means mortals must not assume an afterlife — they must wait and see. But, read figuratively, the poem can be made to yield a sense in which it marks a deep aspect of philosophical investigation. That the company of experiment should be pungent is an idea one can naturally associate with the empiricist tradition; but, taken as a whole, it is the radical empiricism of W. V. O. Quine that the poem seems most perfectly to encapsulate. He argues that a scientific proposition is not tested in isolation but forms an ingredient of a scientific system. If experience turns out unexpectedly, then no one can say in advance which of the system's propositions will collapse. Consequently every proposition, which includes every putative axiom, is in principle corrigible, synthetic. It is true that some axioms, like those of arithmetic, appear unshakeable; their failure is literally unthinkable. But history abounds with cases of 'axiom downfall', where the unthinkable has in time become thinkable. The axiom of parallels is paradigm here, but, as Quine himself points out, even classical logic has come to have its doubters since the advent of quantum theory. And if logic is not immune from revision, then surely nothing is. Axioms are not given an opportunity.

Quine's philosophy underlies some pivotal points of the thesis. So when I attempt to set up a definition of what should count as a prediction, I do not mean to suggest that the result is supposed to capture perfectly the concept of predictability. For one thing, the concept is too multi-faceted to allow such cauterisation. But, more fundamentally, I am conscious that no one can be sure that even this modest definition will in the course of subsequent investigations turn out to be inappropriate. After all this is precisely what I show of a *prima facie* sound definition by Geroch. Experiment, here in the form of *Gedankenexperiment*, escorts us last.

Experiment escorts us in the case of computability too. The prevalent view is that this concept is exactly captured by Turing’s theory, which can thus serve as an axiomatic basis. That confidence in this view runs high is reflected in Hao Wang’s assertion that ‘it is absolutely impossible that anybody who understands the question and knows Turing’s definition should decide for a different concept’.[1] But no one, not even the eminent Wang, can say in advance how computability will appear in the light of new theories and experiments. (Think of the concept of geometry before Einstein and after Einstein.) More to the point is this: the concept of computability is now beginning to take on a radically new appearance, actually in the light of a recent *Gedankenexperiment* in general relativity. To add anything further without the relevant conceptual framework would be unavoidably obfuscatory. So the first chapter helps put the reader in the right frame of mind.

---

[1] Wang 1974, p. 84.

1 Spacetime

1.1 Introduction

The investigations in this thesis are based on classical general relativity (GR). (I keep quantum mechanics switched off throughout.) This chapter provides a sketch of the background theory required for the following two chapters. Being a sketch, there are of course some gaps. For example I will just assume the reader has some acquaintance with the standard exact solutions of the Einstein field equations, e.g. the Schwarzschild, Robertson-Walker and Gödel spacetimes. I will however explain some other, more fundamental notions, including, e.g., lightcones, timelike curves, past and future, slices, and Cauchy surfaces. I will also outline a simple singularity theorem and provide some statements of the cosmic censorship hypothesis. Some other background topics, e.g. Cauchy horizons, Carter-Penrose diagrams, TIPs, will be introduced in the other chapters at the appropriate time. A more than usual emphasis is given throughout to plain language explanations of the mathematics involved. I might add that my preferred text on all these matters is Wald 1984.

1.2 Basic elements

A model in GR consists of a spacetime (M, g_{ab}), i.e. a 4-dimensional Hausdorf, manifold without boundary, which is endowed with a Lorentzian metric.[2] I take the signature of g_{ab} to be (+,+,+,–). The spacetime may also support other fields representing physical phenomena, e.g. electromagnetism, fermions, etc. These matter fields are collectively represented by the *mass-energy* (also called the *energy-momentum*) tensor T_{ab}. The mass (=energy) and curvature are related by the Einstein field equations (EFE):

$$R_{ab} - \frac{1}{2} R g_{ab} + \Lambda g_{ab} = 8\pi \, T_{ab},$$

---

[2] I adopt the abstract index notation; see Wald 1984, p.23.

where R_{ab} is the *Ricci tensor* associated with g_{ab}, $R=R_{ab}g^{ab}$ is the *Ricci scalar*, and Λ is the *cosmological constant*. Hereafter Λ is assumed to be zero unless I indicate otherwise.

The metric g_{ab} assigns lengths to vectors. Being Lorentzian, the lengths may be negative, positive or zero. We say that a vector is *timelike*, *spacelike*, or *null* according to whether its length is positive, negative or zero. Thus in the tangent space at each point of spacetime there is a lightcone structure, with null vectors falling on the cone itself, spacelike vectors falling outside, and timelike vectors falling inside. (It should be emphasised that the lightcone is part of the tangent space of the spacetime and not of spacetime itself.)

Each lightcone has two lobes. To say that spacetime is *temporally orientable* means that it is possible to label in a continuous way one lobe of each cone 'past' and the other 'future'. If the spacetime is not temporally orientable then any distinction between 'past' and 'future' seems to collapse. Most garden variety spacetimes are time-orientable. Indeed, it is easy to prove that all spacetimes based on simply connected manifolds (e.g. ^{4}: so Gödel spacetime is time-orientable) must be time-orientable (Geroch and Horowitz 1979, p. 224). An example of a non-time orientable spacetime is given by Wald (1984, p. 189). I shall hereafter assume time orientability.

A *curve* λ is a continuous map from an interval (or the whole) of ^{1} into M. The image of the interval is formally called a *path*, but for simplicity I will refer to this as a curve also. The *tangent vector* to λ at a point p λ is the vector V V_p that — intuitively — aligns itself with the tangent of λ at p. A *timelike curve* is then a curve with tangent vector everywhere timelike. A *null curve* is defined similarly but with 'timelike' replace by 'null', and a *causal curve* is defined as curve whose tangent vector is nowhere spacelike. A causal curve is *future directed* if each tangent vector points in the future direction.

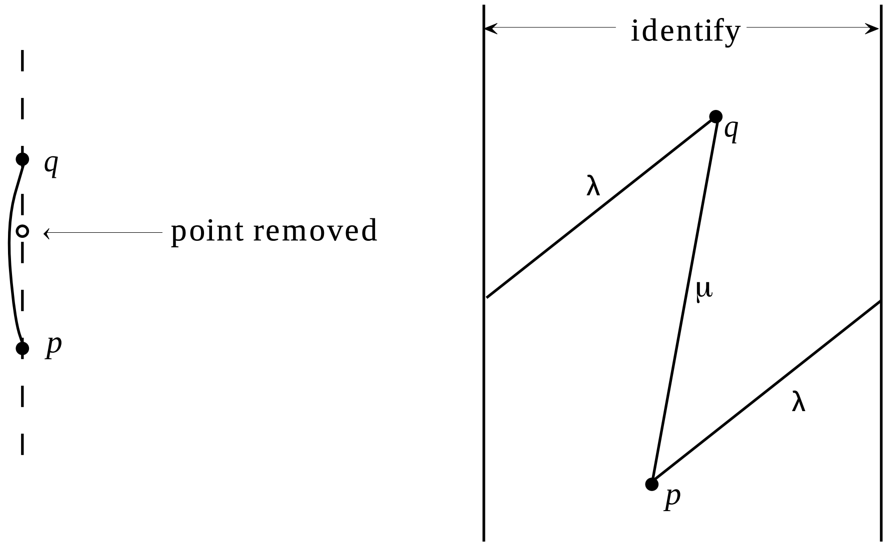
The proper length or proper time τ of a timelike curve is given

$$\tau = \int_{\lambda} (g_{ab}T^a T^b)^{1/2}\, dt,$$

where T^a is the tangent vector to λ and t is a parameter along λ.

The length of a null curve is always zero. The length of a spacelike curve is defined as above, except the sign inside the square root of the integrand is changed from plus to minus. The length of curves which are part timelike and part spacelike is not defined.

A causal curve λ is said to be a *geodesic* if λ is locally of maximal length (i.e. about every point p λ there is a sufficiently small neighbourhood O of p such that the length of λ in O is a maximal). A spacelike curve is said to be a *geodesic* if it is locally of minimum length. Notice that the existence of a causal curve between two points p and q does not ensure the existence of a geodesic between p and q. See figure 1(a). And although geodesics are *locally* curves of maximal length, they need not represent the longest route between two points. See figure 1(b).



(a) (b)

Figure 1. (a) is Minkowski spacetime with a point removed from the line joining p to q. The points p and q can be joined by a timelike curve, but not a timelike geodesic. (b) is a spatially finite segment of Minkowski spacetime with the edges identified. λ and μ are both timelike geodesics joining p to q, but λ is actually shorter than μ.

The spacetime in figure 1(a) illustrates a useful technique for creating a new spacetime from an old one. The general technique is to remove a closed set C of points (or just one point) from a spacetime (M, g_{ab}) to define a new spacetime $(M–C, g^*_{ab})$, where g^*_{ab} is the restriction of g_{ab} on $M–C$. (Notice that C must be closed to ensure that $M–C$ is a manifold without boundary.) Another general technique is illustrated in figure 1(b). Start with a spacetime (M, g_{ab}) with a boundary consisting of two disjoint boundaries S_1 and S_2, and identify S_1 and S_2 so that the metric is smooth across the join.

Although the spacetimes that result from this kind of cutting and pasting are usually not physical, they are useful for providing counter-examples to false claims that on the face of it appear true (these abound, as we shall see).

The lightcone has physical significance, for it represents an absolute speed barrier in this sense: massive bodies are constrained to move on timelike curves (their velocity vector cannot be spacelike or null). Furthermore it is a fundamental tenet of GR that massive bodies which are not affected by any external forces — i.e. are 'freely falling' — move on timelike geodesics. (This can actually be shown to be a consequence of the EFE.) Massless particles (e.g. photons), on the other hand, are constrained to move on null curves. Consequently nothing can travel faster than light.[3] If causal influences are assumed to move on causal curves, then it follows that the lightcones determine the causal structure of spacetime. I will now outline some definitions and results that help characterise this structure, starting with the two most basic definitions: past and future.

1.3 Causal structure

In a spacetime (M, g_{ab}) the *timelike future* of a point $p \in M$, denoted $I^+(p)$, is defined by

$$I^+(p)=\{q \in M \text{ such that } p \text{ can be joined to } q \text{ by a future directed timelike curve.}\}.$$

The definition of $I^-(p)$ is acquired by replacing the word 'future' with 'past'. (In general there will be a self-evident 'temporally opposite' definition obtained by interchanging 'past' and 'future'. The other version will not normally be stated.)

The *causal future* of a point $p \in M$, denoted $J^+(p)$, is defined as $I^+(p)$ except the word 'timelike' is replaced by 'causal'.

Although these sets are defined to apply to single point p, they can each be generalised to apply to a set of points S in the obvious way, e.g.

$$I^+(S)=\bigcup_{p \in S} I^+(p).$$

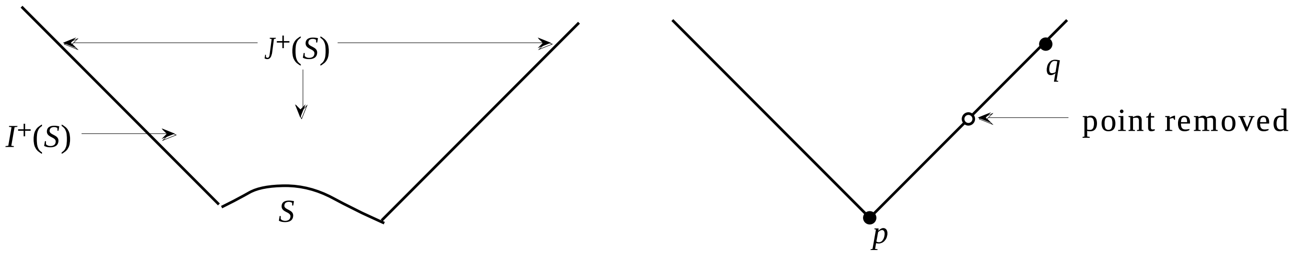
See figure 2(a).

It is a fundamental result that $I^+(S)$ is always an open set (Wald 1984, p. 190). However it is wrong to think that $J^+(S)$ must be closed: a counter-example is

---

[3] I ignore tachyons.

provided by Minkowski spacetime with a point removed, as depicted in figure 2(b). We see that p cannot be connected to q by a causal curve, so $q \quad J^+(p)$. But $q \quad clos\{J^+(p)\}$.[4] Hence $J^+(p)$ is not closed.



(a) (b)

Figure 2.

Every spacetime (M, g_{ab}) is locally Minkowskian in the sense that the tangent space at each point is a Minkowski vector space and also in this sense: about each point $p \quad M$, there is a *convex normal neighbourhood* U of p, i.e. an open set U with $p \quad U$ and such that for $q, r \quad U$ there is a unique geodesic that joins q and r and remains in U. Moreover, $I^+(p) \cap U$ (respectively $\partial I^+(p) \cap U$) is just those points reachable from p by a future directed timelike (respectively null) geodesic that remains in U. See Hawking and Ellis 1973, proposition 4.5.1.

A host of simple results flow from the definitions of J^+ and I^+. One basic lemma is this: if $q \quad I^+(p)$ and $r \quad I^+(q)$, then $r \quad I^+(p)$. In words, 'if q is later than p and r is later than q, then r is later than p'. A less trivial result is: if $q \quad J^+(p) – I^+(p)$, then any causal curve joining p to q must a null geodesic. *Proof.* Suppose a causal curve λ joins p to q. First cover λ with convex normal neighbourhoods, as just defined. Now because the segment of λ between p and q is compact, we can extract a finite subcover. If λ is only partly timelike in any such neighbourhood, then by the previous paragraph λ can be deformed into a wholly timelike curve in that neighbourhood. Continuing through each neighbourhood gives a timelike curve from p to q.

It is crucial to be able to distinguish formally the case of a causal curve which 'goes to infinity' or 'runs into a singularity' from the case in which a curve could be extended. Let λ be a future directed causal curve. Then we say that $p \quad M$ is a *future endpoint* of λ if for every neighbourhood O of p there exists a t_0 such that $\lambda(t) \quad O$ for all $t > t_0$. (Notice that a curve need not contain its endpoints; e.g. in 2-dimensional Minkowski spacetime, take the curve characterised by the set $\{(x,t)½x=0,\ 0<t<1\}$.

---

[4] Recall that for any set A, $clos(A)$ denotes the closure of A, $int(A)$ denotes the interior of A, and ∂A denotes the boundary of A.

Then (0,0) is a past endpoint and (0,1) is a future endpoint.) A causal curve λ is *future endless* (also called *future inextendible*) if λ has no future endpoint. This provides the appropriate distinction: a future endless curve somehow 'runs out of future spacetime', whereas a curve with a future endpoint could continue into the future.

If a timelike curve can be thought of as 'all time at one point of space', then a slice is the dual notion of 'all space at one time'. We first define a set A as *achronal* if no two points on A can be joined by a timelike curve. A *slice* S is then an achronal submanifold of the manifold M. ($M–S$ is therefore open.) Equivalently, a slice S can be defined as a closed, edgeless, achronal set S, where the edge of S is defined as the set of points p such that every open neighbourhood O of p contains a point q $I^+(p)$, a point r $I^-(p)$ and a timelike curve λ from r to q which does not intersect S.

Intuitively: if a set S has an edge, then a timelike curve can 'get from above S to below S without intersecting S'. Thus the edglessness of slices means they 'extend out spatially as far as possible'.

I note some facts about slices. First, if a spacetime can be *partitioned* or *foliated* by one family of slices then it can be foliated by many such families. As an example, take Minkowski spacetime with the canonical coordinate system. This can be foliated by the slices $t=0$ and by the slices $3x–t=0$. Secondly, in general the topology of the slices in one foliation can differ from the topology of the slices in the another (Geroch and Horowitz 1979, p. 245). Finally, some spacetimes do not permit any such foliation, or even a single slice. Think of Gödel spacetime: in this arena there can be no slices because every point can be connected to itself by a timelike curve.

The set $J^+(S)$ can be thought of as the set of events that can be influenced by S. We now seek the set of events that are *determined* by S. For reasons given in chapter 2, the appropriate notion is that of the *future domain of dependence* of a set S, denoted $D^+(S)$, and defined by

$$D^+(S)=\{p \quad M, \text{ such that every past endless causal curve through } p \text{ intersects } S\}.$$

$D^-(S)$ is defined analogously. The union of $D^+(S)$ and $D^-(S)$ is the *domain of dependence* of S, denoted $D(S)$. See figure 3.

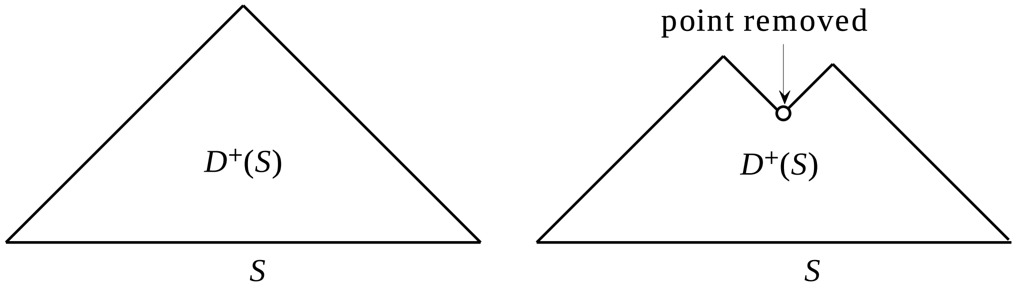


Figure 3. Minkowski spacetime and Minkowski spacetime with a point removed. Since past endless causal curves can enter the hole without registering on S, the hole 'casts a shadow'.

If a spacetime (M, g_{ab}) possesses a slice S such that M=D(S), then we call S a *Cauchy surface* and we say that (M, g_{ab}) is *globally hyperbolic*. In this case, every endless causal curve must intersect S. Examples of spacetimes that admit a Cauchy surface are those associated with the names of Minkowski, Robertson-Walker, Taub, and Schwarzschild. Examples that do not include admit a Cauchy surface: Gödel (because, recall, no slices), Taub-NUT (because timelike curve in Taub region winds round and round in a compact region, and never reaches NUT region), certain of the Kerr (because of a naked singularity — definition coming up shortly), and the pierced spacetime in figure 3. Globally hyperbolic spacetimes have a number of nice properties. For example, if S M is a Cauchy surface then there is always a unique geodesic connecting S to any given point in M (the spacetime is figure 2(b) is not globally hyperbolic). Also, the topology of M is given by $\Sigma\times{}^1$, where Σ is the topology of a Cauchy surface. Because of this last result, one might say that all globally hyperbolic spacetime are alike, but non-globally spacetimes are subtle after their own fashion. Finally, I note that a spacetime (M, g_{ab}) is globally hyperbolic if and only if for any two points x, y M, $J^-(x)\cap J^+(y)$ is either compact or empty (Wald 1984, theorem 8.3.10).

Some spacetimes in GR possess 'causality violations'. Perhaps the simplest example is afforded by identifying the edges of a temporally finite strip of Minkowski spacetime, as depicted in figure 4. This spacetime possesses a *closed timelike curve* (CTC), that is a timelike curve that intersects itself. Indeed every point lies on a CTC. In principle an observer in this world can travel from a point p forward in time until eventually she reaches the point p again. The past and future are well-defined here (the spacetime is time-orientable), but identical. I discuss this curiosity in chapter 2, section 2.8. There are various inequivalent ways to rule out causality violating spacetimes. Imposing the *chronology condition* means that spacetime does not possess CTCs. Imposing *strong causality* means that no causal curve can eventually return arbitrarily close to itself. Strong causality therefore implies the chronology condition, but the converse is false (Wald 1984, p. 197). Then there is *stable causality*, which holds if spacetime admits no closed timelike curves even when the light cones are perturbed slightly (as they presumably are when quantum mechanics is switched on). Stable causality implies strong causality (*ibid.*, p. 199), but the converse is false (Hawking and Ellis 1973, p. 197). Later, at the appropriate junctures, I will state which, if any, of these causality conditions is assumed to hold.

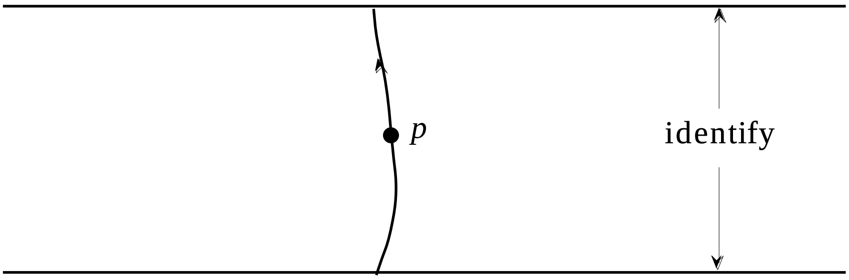


Figure 4. A temporally finite strip of Minkowski spacetime with the edges identified. An observer at p moves forward in time but eventually returns to p.

1.4 Singularities

And the end and the beginning were always there
Before the beginning and after the end.

T. S. ELIOT, *Burnt Norton*

General relativity makes two remarkable and not unrelated predictions: that the universe is expanding and that it possesses singularities. The theory promotes a picture in which the universe (including spacetime itself) began some 10 to 20 billion years ago in a hot and highly dense state of tiny spatial dimensions. This extraordinary or *singular* 'initial' state is known as the big-bang singularity. After the big-bang, the spatial dimensions of the universe expanded and have continued to expand ever since. Whether this expansion will continue indefinitely or reverse into contraction is a matter of current debate (see Hawking and Penrose 1995, chapter 2). In a sense either way is bleak. For eventual contraction implies the universe will come an end (in the 'big crunch' singularity), and continual expansion, though it implies eternal existence, brings with it the promise of a 'heat death' as the available energy tends to zero. The eventual future is at best dead, at worst nothing. (Some would say at worst dead, at best nothing.)

It is usual to call a spacetime with a singularity, *singular*, and a spacetime with no singularities, *non-singular*. Some spacetimes are clearly singular, e.g. the Robertson-Walker big-bang models and the Schwarzschild black hole, and some are clearly non-singular, e.g. Minkowski spacetime. But despite this apparent ease of classification, no definition has yet been advanced which exactly captures what a singularity is — and this despite much research in the last three decades to remedy the problem (see Geroch 1968, Earman 1995, chapter 2). We are consequently left to work with a number of definitions, each suited to particular contexts. The focus in our discussion is on the *existence* of singularities, and within that context the relevant definition is characterised by the slogan: a spacetime is singular just if it is timelike or null-geodesically incomplete. (An *incomplete* geodesic is a past or future endless geodesic of finite length.) By equating a 'singularity' with one of its effects on spacetime, this approach neatly side-steps the vexed question of what a singularity actually is.[5] The physical picture corresponding to this mathematical definition is one in which a photon or freely falling particle can, so to speak, disappear prematurely off the edge of spacetime.

(Geodesic incompleteness is arguably neither a necessary nor a sufficient condition for the existence of a singularity. It is not sufficient because there are examples of spacetimes which are geodesically complete but which possess a future endless timelike curve with bounded acceleration and finite length (Geroch 1968). Physically this means an astronaut with a finite amount of fuel could disappear off the edge of spacetime in a finite time. The situation has a decidedly singular ring. On the other hand, the condition is apparently not necessary either, for there are geodesically incomplete *compact* spacetimes. The point here is that arguably a compact spacetime is *necessarily* singularity free because it contains all its limit points, i.e. there is no 'singular edge'.)

What is known about the existence of singularities derives primarily from a series of remarkable results developed in the late 1960s by Geroch, Hawking, and Penrose. These so-called singularity theorems show that singularities are expected in a wide range of physically reasonable models (again, recall the Robertson-Walker big-bang models and the Schwarzschild black hole: each of these spacetimes is geodesically incomplete). Each of the theorems is actually a purely geometrical result with essentially the following form. There is a 'local' condition which says nearby geodesics are irrevocably drawn closer together and eventually cross; and there is a global condition, e.g. the spacetime possesses a Cauchy surface or compact slice. The combination of these two conditions is shown to be incompatible with geodesic completeness. What gives the theorems physical significance is that these geometric

---

[5] An analogy is found in the case of schizophrenia. This illness is not defined in neurophysiological terms because no neuropathology has yet been identified. Instead it is defined in terms of its *symptoms*, e.g. hearing voices, paranoia.

conditions may obtain in our universe. Here I sketch the proof of a particularly simple (and weak) singularity theorem.

Consider a congruence of timelike geodesics which emanate normally from a slice S. Let θ be the average expansion of nearby geodesics. The rate at which θ changes along the congruence is given by Raychaudhuri's equation (Wald 1984, p. 218), which can be stated as

$$\frac{d\theta}{d\tau} = -\frac{1}{3}\theta^2 - R_{cd}\,\xi^c\xi^d - \sigma_{ab}\sigma^{ab},$$

where ξ^a is a tangent vector to the congruence and σ_{ab} measures the *shear* of the congruence.

The $R_{cd}\,\xi^c\xi^d$ term can be rewritten in terms of the energy-momentum tensor by appeal to the EFE (this is where the physics first enters).

$$R_{cd}\,\xi^c\xi^d = 8\pi\,[T_{cd} - \frac{1}{2}\,T\,g_{cd}]\,\xi^c\xi^d = 8\pi\,[T_{cd}\xi^c\xi^d + \frac{1}{2}T],$$

where $T = trace\,(T_{ab})$.

The right hand side is believed to be positive for all physically reasonable forms of matter. If this is true, then the *strong energy condition* is said to hold. In this case, and noting that $\sigma_{ab}\sigma^{ab}$ is manifestly positive, the Raychaudhuri equation yields

$$\frac{d\theta}{d\tau} + \frac{1}{3}\theta^2 \leq 0,$$

which implies

$$\frac{d}{d\tau}(\theta^{-1}) \geq \frac{1}{3},$$

and so

$$\theta^{-1}(\tau) \geq \theta_0^{-1} + \frac{1}{3}\tau,$$

where θ_0 is the initial value of θ.

If θ_0 is negative, i.e. the congruence is initially converging, then $\theta \to -\infty$ within a proper time $\tau \leq 3/|\theta_0|$. This implies that the geodesics of the congruence have begun to cross each other.

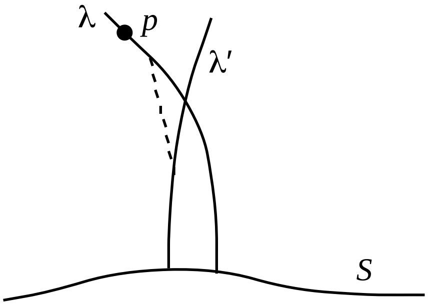


Figure 5. Rounding off the corner creates a longer curve from S to p.

This does not by itself imply geodesic incompleteness (counterexample: take S to be the hyperboloid in Minkowski spacetime given by $t=(r^2 + 1)^{1/2}$.); for that we require an extra condition, namely, that S is a Cauchy surface. To see how this works, suppose for the sake of a contradiction that (M, g_{ab}) is timelike geodesically complete. Now follow a geodesic from S for a distance that exceeds $3/|\theta_0|$, and label the point p, as in figure 5. Now since (M, g_{ab}) is globally hyperbolic, there is a timelike curve of maximal length from S to p. But we know from the previous paragraph that λ must be crossed by a nearby geodesic λ' before p is reached. Thus λ cannot be maximal, because rounding off the corner at the intersection of λ and λ' produces a curve from S to p of length greater than λ. This contradicts the initial assumption that (M, g_{ab}) is timelike geodesically complete.

In short: if the universe is globally hyperbolic, is everywhere expanding (or indeed contracting), and the strong energy condition holds, then there will be timelike geodesic incompleteness (a singularity). This is a not insignificant result, but geodesic incompleteness can actually be shown to follow from much more plausible (i.e. weaker) assumptions. Specifically we have the Hawking and Penrose (1970) singularity theorem (chapter 2, section 2.9) which reveals essentially the following. Closed universes (i.e. those admitting a compact slice) are generally singular. On the other hand, for open universes, there are generic ones which are singular and generic ones which are not. It seems that an extra ingredient is required, e.g. a closed trapped surface (which occurs in black-holes) or a slice with convergence bounded away from zero (contraction everywhere or expansion everywhere), before one can infer that an open universe is singular.

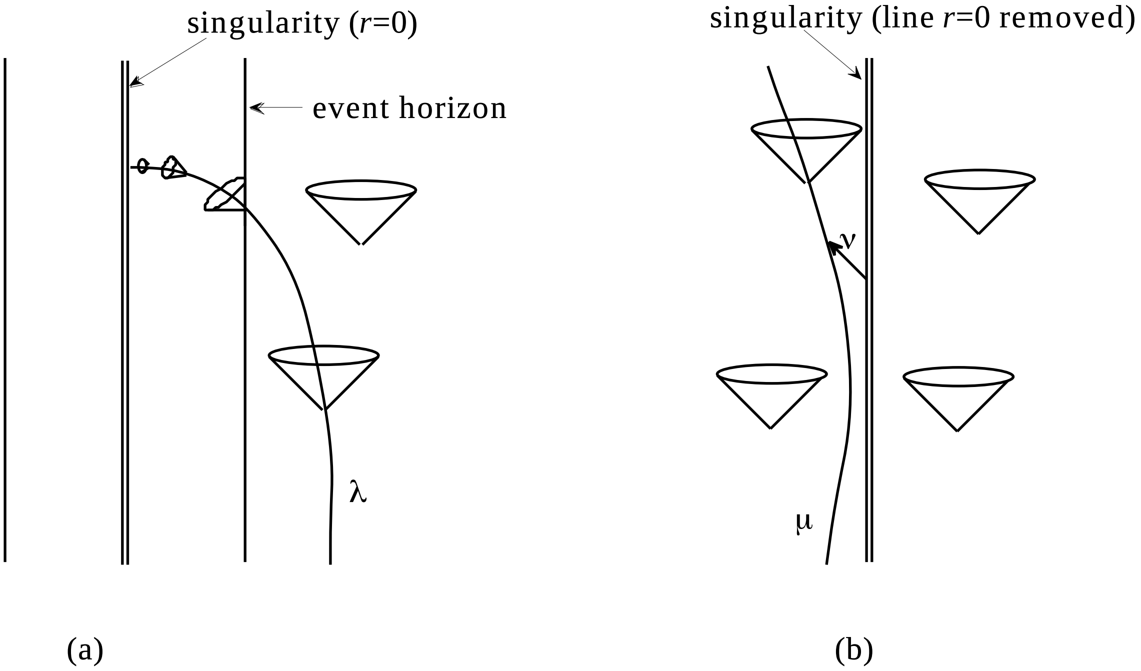


Figure 6. (a) shows part of the Schwarzschild black hole solution. The singularity is not directly experienced by the observer λ. In (b) a naked singularity is formed by removing a line from Minkowski spacetime. A photon like ν emerges from the singularity and arrives at μ without any warning.

1.5 Naked singularities and cosmic censorship

We expect singularities, but what kind of singularities do we expect? The very word ‘singularity’ suggests something untoward, but some singularities are decidedly worse than others. On the tame side consider the singularity in the Schwarzschild solution. I say ‘tame’ because although this singularity will necessarily snuff out any observer who is fool enough to cross the event horizon, it will not cause trouble for observers in the larger region beyond. Indeed, an inspection of the light cone structure reveals that even those observers who do cross the event horizon cannot possibly *experience* the singularity’s bite, for they are terminated at ‘the moment’ when that experience would become possible.[6] See figure 6(a). In this sense, the singularity, and *a fortiori* any attendant bad behaviour (lawlessness) at the singularity, is completely hidden from all.

---

[6] Since the singularity is not itself part of spacetime, there is no moment of death (even though the observer’s life is cut short). According to Wittgenstein (1921): ‘Death is not an event in life: we do not live to experience death’ (*Tractatus*: 6.4311). But death by falling into a black-hole is not an event *at all*.

If the Schwarzschild singularity is deemed tame, then the singularity depicted in figure 6(b) is surely wild. In this spacetime, an observer (e.g. μ) can not only experience the singular region, but can do so at arbitrarily close quarters. This kind of singularity is accordingly labelled 'naked' (formal definition coming up). The danger posed by nakedness is that some events in spacetime will fail to be determined by *any* amount of past data. The future will be full of surprises.

Do naked singularities exist? The *cosmic censorship hypothesis* (CCH) is precisely the claim that the answer is no. Proposed by Penrose in 1969, the CCH has emerged as the most important unsolved problem in classical general relativity. (One might imagine that the singularity theorems would help settle the issue, but they prove *only* the existence of geodesic incompleteness; questions regarding the nature of the corresponding singularity, e.g. whether it is in the past or the future, or whether or not there is an event horizon, are untouched.) The principal argument in favour of the CCH derives from the models of gravitational collapse. Exact spherically symmetric collapse is represented by the Schwarzschild black hole solution. The results of performing small perturbations on this solution suggest that small departures from spherical symmetry will also produce a black hole rather than a naked singularity. (There must still be a singularity, by the singularity theorems.) Moreover, in the case of gravitational collapse with charge and/or rotation, as represented by the Kerr solutions, the general picture seems to be that the black hole solutions are stable and the naked singular solutions are unstable. Taken as a general statement about gravitational collapse this amounts to one way to frame roughly the CCH.

Unfortunately the roughness here is not easily smoothed. Indeed one of the major difficulties with deciding whether or not the CCH is true is that as yet there is no entirely satisfactory framing of its statement (Earman 1995, chapter 2). In what follows I outline three important attempts.

Strong Cosmic Censorship (*SCC*)

Roughly: spacetime should be deterministic in the sense that it all evolves from a single slice.

Definition 1.5.1 A spacetime (M, g_{ab}) satisfies strong cosmic censorship just if (M, g_{ab}) possesses a Cauchy surface (i.e. if (M, g_{ab}) is globally hyperbolic).

Since any globally hyperbolic spacetime can be foliated by a family of Cauchy surfaces, the picture of a spacetime satisfying SCC bears some resemblance to

Newtonian spacetime with its foliation of absolute timeslices. Of course, because of the variable lightcone structure, the slices in the relativistic case may be forced to bend round singularities, but the singularities themselves are always stationed either ‘before’ the first slice or ‘after’ the last. The singularities are therefore of two kinds: the visible, unvisitable (e.g. the big-bang), and the invisible, visitable (e.g. the big-crunch, black holes).

Weak Cosmic Censorship (WCC).

Roughly: all singularities must be hidden from view. Put another way: the region visible from future infinity is singularity free. ‘Future infinity’ can be given meaning only in *asymptotically flat* spacetimes (see Wald 1984, chapter 11), where it is defined as $\ ^+$, i.e. the set of points adjoined to spacetime at future null infinity. (The interior of a black hole is then defined as $M – J^-(\ ^+)$.) We have:

Definition 1.5.2 An asymptotically flat spacetime (M, g_{ab}) satisfies weak cosmic censorship just if $(J^-(\ ^+), g_{ab})$ is globally hyperbolic (i.e. $(J^-(\ ^+), g_{ab})$ satisfies strong cosmic censorship)

(As one would expect WCC does not imply SCC (Earman 1995, p. 73). But in fact SCC does not imply WCC (Penrose 1978, p. 234). WCC and SCC are therefore logically independent, and so the terms ‘weak’ and ‘strong’ are not entirely appropriate.)

No naked singularities

Roughly: no observer can witness another observer disappearing off the edge of spacetime. Following Geroch and Horowitz (1979, p. 274), consider the following schema for defining for a given spacetime (M, g_{ab}) the set of nakedly singular points, N, by

N={$p \ \ M$: $I^-(p)$ contains a future endless timelike curve λ which is ______}.

That there is a cosmic censor is equivalent to the statement that N is empty, but different versions of the CCH are produced by inserting different conditions into the blank. Leaving the blank blank gives the strongest version, which is actually equivalent to SCC (see the corollary of proposition 3.5.1). Alternatively, inserting ‘a geodesic’ focuses attention on singularities reachable by freely falling observers.

But it is not an aim of this thesis to discuss the vexed question of the truth value of any of these forms of CCH (see Earman 1995, chapter 2; Hawking and Penrose

1995, chapter 2). Rather my focus will be to show how issues about censorship are tied up with predictability and computability. I begin now with an examination of the former.

2 Predictability

If you can look into the seeds of time,
And say which grain will grow and which will not,
Speak then to me...

Macbeth 1.*iii*.58

2.1 Introduction

Prediction is an obvious feature of everyday science, but what would actually be required so that an observer could perform a prediction that was bound to be true? This is the central question of this chapter. I will show that the physical fields can be conducive to prediction, but that by contrast the geometry of spacetime often creates insurmountable obstacles (section 2.1). This leads to the definition of a 'locally predictable' (LP) spacetime, i.e. a spacetime that allow at least one prediction (section 2.2). The structure of LP spacetimes is investigated in section 2.3. Then a variety of generalised LP spacetimes are defined in sections 2.4 and 2.5, which is followed in section 2.6 by some examples. In section 2.7 I investigate the problem of predicting not just the physical fields but spacetime itself, and show how this notion is connected with the previous results. A worry is created in section 2.8 that LP spacetimes yield a logical paradox; but the worry is soothed. In the final section before the summary I look at the subtle relationships that exist between singularities and predictability.

I begin though by examining a concept which in some sense underpins prediction — determinism.

Determinism

Conceptions of determinism go back at least as far Aristotle (recall the sea battle), but I will launch the discussion from Laplace, who in 1819 wrote:

'An intelligence knowing all the forces acting in nature at a given instant, as well as the instantaneous 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 the past would be present to its eyes.’ (Quoted in Popper (1982, p. xx).)

But modern interpretations of determinism eschew the epistemic aspect of Laplace’s idea in favour of something wholly ontological. We now say that a theory is deterministic if the future and the past of the world are fixed by the laws of the theory after an instantaneous state of the world has been given. Let us sketch this idea more fully.

Associated with any theory T is a set consisting of models all of which obey the laws of T. We can think of these models as being representative of *worlds* that are physically possible (according to the theory T). Let the set of such worlds be denoted W_T, where the T is to remind us that physical possibility is fixed by what T allows. Then we say that a theory T is deterministic if, given any pair of worlds w, w' W_T such that w matches w' at some time t_0, then w matches w' for all time. That is to say, only one world is compatible with one instantaneous snap-shot of the world.

Determinism thus defined makes no reference to anyone’s or anything’s epistemic powers. Notice also that determinism is not so much a feature of the world as a feature of theories of the world or, in picturesque terms, a feature of physically possible worlds. To say the world is (non-)deterministic is a confusion: only theories can be (non-)deterministic.

Now according to folklore, determinism holds in theories based on Newtonian spacetime. For example, one hears it said that within the context of Newtonian gravity the momentary position and velocity states of the sun and planets determines the future and past states of the solar system. This echoes Laplace’s sentiment.

But in fact this folklore is misguided. Take the theory just mentioned, or, more precisely, take the theory T to be Newtonian gravity operating on point particles with fixed mass in Newtonian spacetime. Then T is not deterministic with respect to a timeslice (=all space at one instant of time) S because it is possible according to T for a particle to enter spacetime without registering itself on S. S, in other words, does not determine the particle configuration throughout the world. Such particles Earman (1986, chapter 3) wittily calls *space-invaders*. Their existence in T follows because there are initial configurations of four rocket ships that evolve so that a mass moving between them (without collisions) approaches infinite speed within a finite time (proved by Gerver (1986), but see also Mather and McGehee’s (1975) similar result which relies on collisions). The space-invader is the result of taking the temporal reflection (which is an allowed state of affairs according to *T*) of this tearaway mass.

Adjoining conditions to *T* designed to outlaw space-invaders turns out to be both *ad hoc* and/or technically problematic, as Earman himself points out. It therefore appears that any 'natural' theory of Newtonian gravity is indeterministic. Moreover, since the source of the problem really lies with Newtonian spacetime — that it possesses no intrinsic means of limiting the speed of particles — space-invaders are expected as a generic feature of Newtonian theories. And wherever theoretical space-invaders strike, determinism is bound to fall.

Fortunately the role of determinism in the relativistic realm is generally more assured. To speak about determinism here first requires finding a relativistic determining region analogous to the Newtonian timeslice. A slice, as defined in the previous chapter, is the obvious choice. Like the Newtonian timeslice, slices are very effective at determining large (4-dimensional) regions of spacetime, as we shall see. Indeed, in the case of globally hyperbolic spacetimes a slice can be found which determines the whole spacetime. Determining regions in the form of achronal sets with edges also play a role in relativistic determinism. These can be thought of as 'part of space at one time', and they determine a region of spacetime correspondingly smaller than that determined by a slice.

In short: we say that a *relativistic theory T is deterministic with respect to* a *region A* (which is often taken to be an achronal set) if the physical states on *A* serve to fix the physical states throughout (M, g_{ab}). And again can put this in terms of models of *T*: a theory *T* is deterministic with respect to a region A if, given any pair of worlds w, w' W_T such that w matches w′ on A, then w matches w'. (This definition is perfectly adequate for our purposes, but I note that defining determinism precisely is actually a matter of subtlety and controversy. This has come to light particularly through the recent investigations into the so-called 'hole argument'; see for example Butterfield (1989), Rynasiewicz (1994), and, for further references, Butterfield, Hogarth, and Belot (1996).)

The paradigm deterministic example is the theory of Maxwellian electrodynamics in Minkowski spacetime. This follows from the fact that the electromagnetic field and source J^μ at each point in spacetime are fixed once the Maxwell tensor and source J^μ has been specified on any slice of the form *t=constant*, where *t* is the usual time coordinate.

I have yet to say what set of events a given achronal set S actually determines. To that end I will make this assumption about the nature of causality: the state at any point p depends only on influences travelling on future directed causal curves with endpoint p. This would be violated by, for example, the existence of tachyons (Earman 1986, p. 75) or, less controversially, by the existence of 'non-local' aspects within quantum mechanics (see Redhead 1987). But within my self-imposed classical limitations, the assumption remains reasonable.

So suppose this causal condition holds in a spacetime (M, g_{ab}), and suppose further that p $D^+(S)$. Then every past endless causal curve through p must intersect S, which means that the data on S should be sufficient to fix the state at p. Thus the region intuitively determined by S contains the future domain of dependence of S. On the other hand, future events outside $D^+(S)$ will not be determined by S because they could be affected by influences that fail to register on S. Consequently $D^+(S)$ is *precisely* the region of spacetime determined by S. (Although this gives the future domain of dependence a physical interpretation, its principal use is as a mathematical construct.)

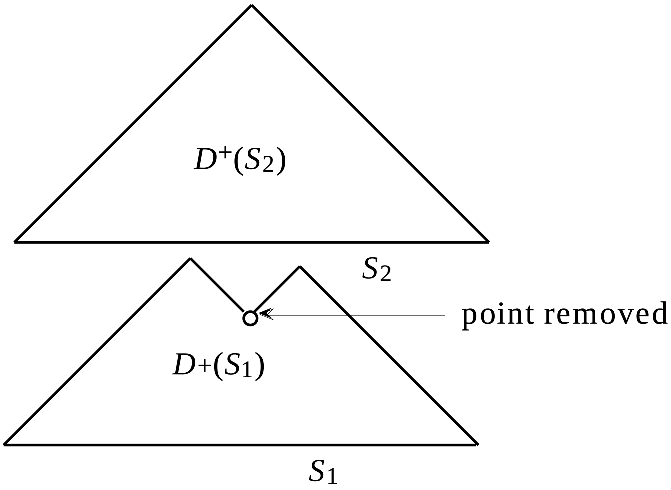


Figure 7. Minkowski spacetime with a point removed. The region $D^+(S_1)$ $D^+(S_2)$ is determined collectively by S_1 and S_2, but it cannot be determined by any single achronal set.

I note that despite the impressions given in most GR textbooks, there is no *a priori* reason why a determining region in a relativistic spacetime must be achronal, anymore than a determining region in Newtonian spacetime must be a timeslice. Indeed in certain contexts it is necessary to choose a non-achronal set simply because there is no achronal set that will do the job. Take the pierced Minkowski spacetime depicted in figure 7, for example. Here the region given by

$D^+(S_1)$ $D^+(S_2)$ is determined by the non-achronal set S_1 S_2, but there is no single achronal set whose future domain of dependence contains $D^+(S_1)$ $D^+(S_2)$. To see this, try drawing an achronal set first above the hole, and then below. Both attempts fail.

The relevance of this apparently trivial point is found in section 2.3. There we encounter a situation in which an event is determined by a specific region but not by any achronal set within that region — and this despite one's intuitions, and at least one contrary claim in the literature.

Wheat and chaff

At the outset it is important to make a distinction between what might be called *true prediction* and *sure prediction*. A true prediction is simply a statement about a future event that is true. Keep guessing that the craps player will throw a six and soon or later you will make a true prediction. And because truth is all that matters here, *knowledge* of the truth value of the prediction — which may occur only later (by observing the thrown dice), and perhaps quite indirectly (by receiving winnings) — is another and quite separate issue. One can make a true prediction without knowing it is true.

Then there is what I will call sure prediction. This is a statement about the future which can be fully justified at the time it is made. In other words, there is every reason to believe that a sure prediction is a true prediction. Naturally the kind of justification I have in mind is to be given in terms of the laws of physics. So, for example, a sure prediction might conceivably be: the falling ball will impact on the floor after 3 seconds because the laws of physics, e.g. Newton's law of gravitation, viscosity laws etc., dictate that it must. On the other hand, a sure prediction of what the craps player will throw may not be possible even in principle, for it may be that the system (a) evolves according to chaotic laws and (b) is subject to (even tiny) random quantum fluctuations. Our forecasts will be ruined because the micro-uncertainty caused by (b) will be hyper-inflated by (a) to macro-uncertainty.

Of course, the 'fully justified' account required for sure prediction cannot be an absolutely indubitable account; that is obviously impossible because our knowledge of physical laws is always to some extent limited. Rather it means the next best thing: 'justified to the best of our current knowledge'. I grant that this is somewhat vague, but it performs well enough to separate the kind of prediction that operates in an idealised science from the kind of prediction that amounts to mere guessing. Real science of course exists between the two.

The kind of prediction under investigation here will be ‘sure prediction’. Admittedly this side-steps a host of other kinds of prediction that occur in science (e.g. predicting quantum measurements, the existence of spacetime singularities, or the weather) and out of science (e.g. predicting shares prices, or love). Yet the manoeuvre is justified, for two reasons. First sure prediction clearly does mirror, at least to some extent, the kind prediction that occurs in normal, everyday science; and second, it can be accurately captured in a mathematical formulation, as we shall shortly see. So by the word ‘prediction’ I will hereafter just mean ‘sure prediction’.

Some conditions for predictability

Observers cannot make predictions without *some* knowledge of the future state of world, for otherwise literally anything could be possible. So in the first instance I will grant them knowledge of:

Conditions 2.1.1.

(a) the laws governing the physical fields;

(b) the global structure of spacetime, or at least the structure upto the event being predicted.

Assumption (b) is reasonable enough if the predicted event lies in the near future, but otherwise it is difficult to accept as a brute fact. However in section 2.7 I will show if the EFE hold then (b) essentially follows from (a) together with the relevant past data. The reason for not employing the EFE right at the outset is that the definitions, propositions and counter-examples relating to prediction can be more readily set-up if one does not have to contend with satisfying the EFE. And, as one would hope, it will turn out that the investigations conducted under assumption (b) will continue to be relevant once (b) is dropped.

How could an event be predicted? The basic idea is this: a spacetime theory T (again *theory* not *world*) allows the prediction of (the fields at) a point p if an observer (call her *Sibyl*) can gather together the data that determines p and establish the state of p before p has happened. (Hereafter phrases like ‘establish the state of p’ will often be shortened to ‘establish p’.) The process can be broken down into three parts.

Conditions 2.1.2

(1) p must be determined by the appropriate past data;

(2) Sibyl must be able to gather that data before p, say at q.

(3) Sibyl must establish p from this data before p, say at q.

(2) is essentially a geometrical condition, which I will deal with in the next section. For now, let us assume that (2) holds (so Sibyl can gather the data by q in the manner illustrated schematically in figure 8), and address conditions (1) and (3), which are concerned with the physical fields on spacetime.

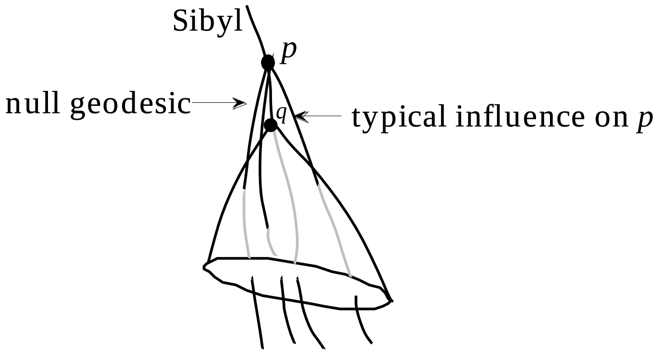


Figure 8. A schematic illustration of how Sibyl could make a prediction. Every influence on p is registered in q's past, which provides the information for Sibyl to calculate the state at p.

The role of the physical fields

Keep an eye on figure 8. Condition (1) is satisfied if the influences on p lying in $J^-(q)$, together with their governing laws, determine the state at p. For otherwise knowledge of the data in $J^-(q)$ (which is all Sibyl at q has to go on) will not be sufficient to infer the state at p.

And as a step towards satisfying (3) we might insist that p's state is a continuos function of the data in $J^-(q)$, because otherwise Sibyl would have to perform the unfeasible task of measuring the physical field with *absolute* accuracy. (It is true that without this accuracy Sibyl with fail to predict p exactly, but if the continuity condition is in place then she can predict p with greater and greater accuracy by making finer and finer measurements of the past physical fields.)

Technically speaking, the combined conditions of the last two paragraphs amount to saying that the theory embodying these laws admits a *well-posed initial value formulation*; that is, that the dynamical evolution of the system is completely determined by and is continuous with respect to the 'appropriate' initial data (see Wald 1984, chapter 10). Many classical theories are of this kind, e.g. Maxwellian electro-magnetism and the classical Klein-Gordon field in Minkowski spacetime.

Now it depends on one's construal of 'in principle', but I am happy to count this well-posed initial value formulation condition as sufficient for prediction *in principle* (of course if (2) holds). Obviously a claim of this kind is not true or false; the best that can be said is that it is an interesting statement to make. The justification in this case is partly a matter of hindsight, because it will transpire that it is the *geometry* of spacetime, and not the physical fields, that plays the crucial role in prediction. Specifically it will be shown that in general the geometry creates an insurmountable obstacle to prediction. The strength of this negative result is therefore enhanced if prediction is given every other opportunity to work, by, for example, making the conditions on the physical fields as weak as is credibly possible. This is why I am content to say the physical fields need only admit a well-posed initial value formulation.

But before turning to the geometric issues, I want to add some comments — comments which are largely incidental to my discussion but important perhaps to those who are dissatisfied by my construal of 'in principle' — about how condition (3) might be met more realistically.

Since Sibyl presumably has to establish the state at p using a *computer*, the state at p should be computable from the $J^{-}(q)$ data. That is to say, a suitably programmed Turing machine could, when given this data, compute the state at p in a finite number of steps.[7] While this assumption may sound innocuous, it is perhaps sobering to note that Pour-El and Richards (1981) have shown that there are solutions of the ordinary wave equation that are deterministic with respect to an initial data set but whose future evolution is not computable. (This, incidentally, highlights the distinction between determinism and computability.) If non-computability did prove to be a generic feature of physical theories, as this result might suggest, then there are those who would claim that predictability has been dealt the *coup de grâce*.[8] This remains to be seen, however, because the trouble-causing data sets found so far are not 'smoothly varying', which is something one would normally require of a physical field.

I might add, parenthetically, that the result by Pour-El and Richards, ingenious as it is, should really come as no surprise. Just because we humans use Turing machines to establish the evolution of physical systems does not imply, or even suggest, that Nature itself evolves in a Turing computable way. It is true that there is an intuitive idea that Nature needs to calculate her evolution, so to speak, 'in advance' . But the

---

[7] By 'computable' I mean here 'Grzegorczyk computable' because the computer is manipulating functions of the reals (not integers); see Earman 1986, p.116.

[8] The non-Turing computers discussed in chapter 3 would not help because they depend on a naked singularity which would itself destroy the possibility of prediction.

idea is not to be taken seriously. Nature can surely proceed without *knowing* where she's going. After all, humans do. (Here again we touch on the determinism / prediction distinction.) And in fact it is not entirely clear to me that even the *notion* that Nature computes her own evolution is a consistent one. The problem is that any such computer must by definition be part of Nature, which raises a question: what computes the evolution computer's evolution? The same computer? Or another? A self-referential paradox seems to loom over the first option, an infinite regress over the second.

Returning to condition (3), we might also want to guarantee that Sibyl has enough time to perform the said computation. To this end it should be ensured that the computational rate required to compute the state at p should not exceed, or at least not wildly exceed, that which is physically possible, or is physically reasonable or is practically possible, or possible using current computers, or ______, where the blank is filled with your favourite kind of possibility.

I could continue to sketch the ideal home for (3), but, as I just remarked, the most interesting aspect of prediction lies elsewhere, in the geometry of spacetime.

The awkward role played by geometry

Attention can justifiably be restricted to spacetimes which do not possess closed timelike curves, because it is a trivial matter to predict an event in one's future if that event also lies in one's past. In fact, it will be convenient to assume the very slightly stronger condition of strong causality. (This global condition is dropped in section 2.9.)

It is helpful to begin by considering how a prediction might be engineered in the simplest of all relativistic spacetimes. So suppose that our observer Sibyl is in Minkowski spacetime and that she wants to make a prediction about the future state of an appropriately well-behaved (in the above sense) physical field, e.g. the Maxwell field. Then this much is clear: if Sibyl were equipped with the data on some Cauchy surface then she could use Maxwell's equations to predict the entire future evolution of the field. However, since Sibyl is an *actual* observer and not some disembodied spirit who is just 'handed' the Cauchy surface data (recall Laplace's demon), this data can only be revealed to Sibyl after it has passed into her past light cone — which is impossible in Minkowski spacetime because no point lies to the causal future of a Cauchy surface. See figure 9. (According to lemma 2.5.3 a Cauchy surface to the past of a point is necessarily compact; but the Cauchy surfaces of Minkowski spacetime are all $\Re^3$.)

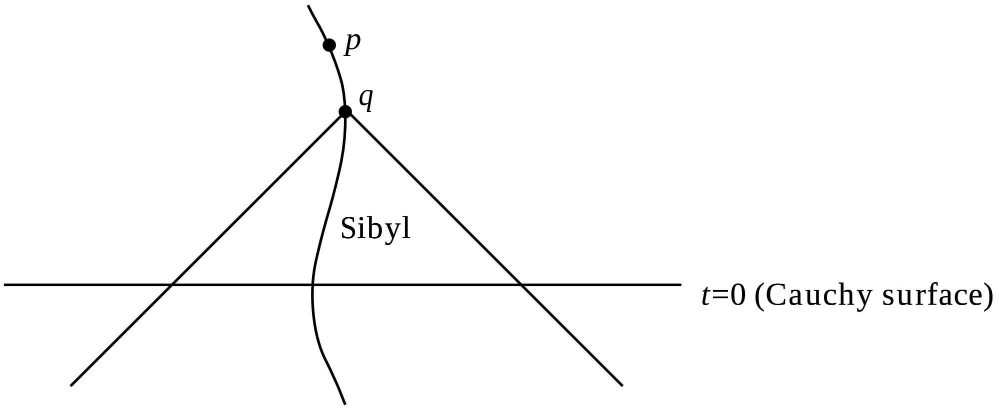


Figure 9. Minkowski spacetime. The event p is determined by the Cauchy surface t=0, but Sibyl (currently at q) cannot position herself to the future of this (or any) Cauchy surface.

Undaunted, perhaps Sibyl could try to perform the less ambitious task of predicting just one future event. This would mean only having to examine an appropriate determining region S of *finite extent* ($p \quad D^+(S)$) — a task which is perfectly possible. But alas prediction is again thwarted, since one can show that by the time the determining region has passed into her past light cone, ready for examination, Sibyl has passed into the future of the event she wanted to predict. Thus located, she can only *retro*dict (see Geroch 1977, p. 88; Earman 1986, p. 193). See figure 10. (A corollary of proposition 2.3.2 below shows formally that prediction in Minkowski spacetime is impossible.)

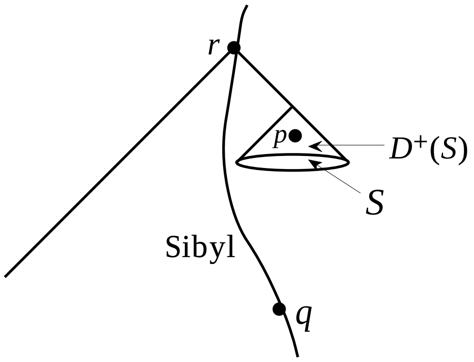


Figure 10. Minkowski spacetime. Since $p \quad D^+(S)$, S determines p. At q Sibyl decides to engineer a prediction of p by reading the relevant data off S. But to do this she has wait until r when S becomes accessible, and unfortunately by that time p is in her past. Failed again.

2.2 Critique of Geroch's paper 'Prediction in General Relativity'

The first rigorous definition of a 'predictable event' was given by Geroch (1977).[9] Unfortunately the definition is not entirely satisfactory. In fact, although the paper has been justly influential, it actually carries a number of important technical errors. One of my aims here is to provide the relevant corrections. The salient points of Geroch's paper are as follows.

> (G1) A precise definition of a predictable event.
>
> (G2) An unproved claim that every LP spacetime admits a compact slice.
>
> (G3) An unproved claim that the data determining a predicted event can always be read off an achronal set.

I will deal with (G1) now, and then (G2) and (G3) in the next section.

Addressing (*G*1)

According to Geroch, we should say that an event p is predictable from an event q if:

(a) every past endless causal curve through p intersects $I^-(q)$;

(b) $I^-(p) \not\subset I^-(q)$.

He explains his choice of conditions as follows 'The point q represents the point (of our predicting observer) at which all the information has been collected. Then the set $I^-(q)$ represents that region of spacetime from which information could reach q. The first condition [(a)] requires, physically, that every signal that could affect p must have come from $I^-(q)$, i.e. that every such signal could have been recorded and carried to q. The second condition [(b)] requires essentially that p is not in $I^-(q)$, i.e., that we have prediction at p rather than retrodiction.'

My first point of criticism is a trivial one: condition (b), as stated by Geroch, is just an obtuse way of stating that $p \quad I^-(q)$.

---

[9] Budic and Sachs' (1976) offer an alternative definition, as we shall see, but unlike Geroch they do little to connect it with the concept of predictability; rather it is presented more as an interesting mathematical definition.

The second point though is more significant. His condition (a) does not seem to reflect his corresponding physical idea because the region from which information can reach q is surely $J^-(q)$, not $I^-(q)$.

Other shortcomings will become clear in the light of an example which contains events that are predictable according to Geroch's (a) and (b). I will speak of these as *G-predictable* events.

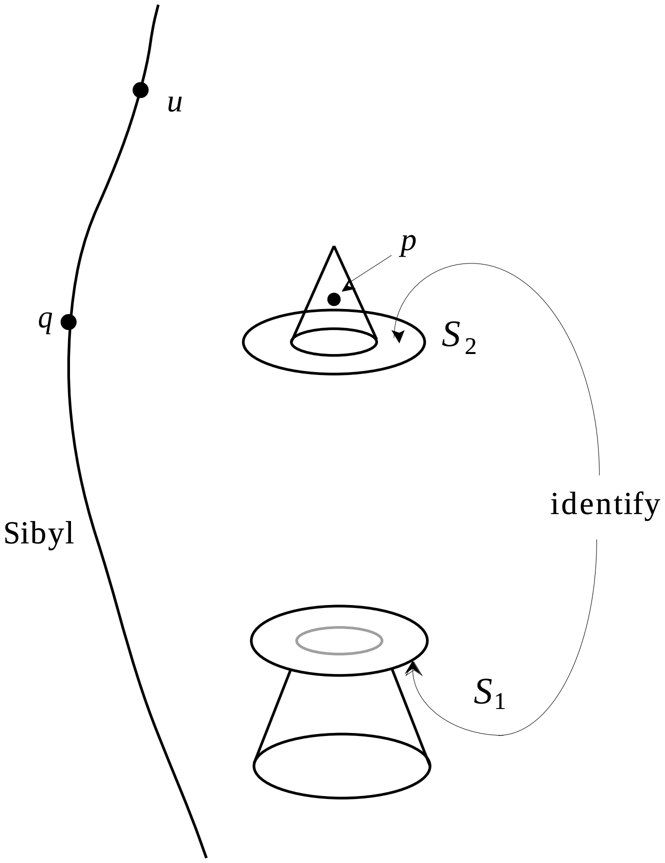


Figure 11. A four-dimensional spacetime obtained by drawing two small, spacelike, three-dimensional disks in Minkowski spacetime and then identifying the 'lower' edge of S_1 with the 'upper' side of S_2. The 'rim' of each identified disks is removed (because otherwise the manifold would have a boundary).

(The construction in figure 11 is a spacetime because the metric joins across the disks smoothly. To gain a mental picture, it helps to imagine that top of S_1 and the bottom of S_2 are joined by a smooth tube or 'worm hole'.)

In the spacetime depicted in figure 11 the point p is G-predictable from q. This follows because every past endless causal curve through p intersects $J^-(q)$, and $q \quad I^+(p)$.

Now Geroch's condition (b) ensures that q is not in p's future but it does not ensure that q is in p's past — the spacetime in figure 11 illustrates that much. But this suggests that condition (b) is inappropriate, for surely the 'pre' suffix of the word 'prediction' denotes 'before' or 'in the past', not 'not in the future'. It is true that in Newtonian spacetime there is no distinction ('present' aside) between 'in the past' and 'not in the future', but that does not give us license to conflate these relations in the relativistic realm. With the distinction rightly upheld, relativistic prediction means the inference of a future event from a past event. Of course this is not to say that inference of events that are not in the past is uninteresting; only that it does not deserve the appellation 'prediction'. It would make more sense to call G-prediction 'inference of non-past events'.

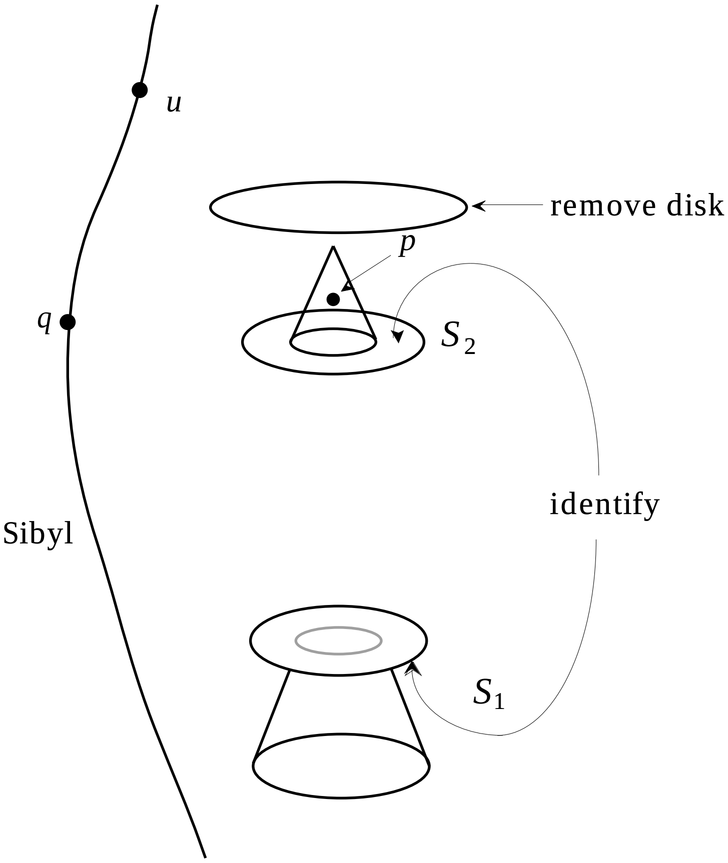


Figure 12. A spacetime obtained by taking the spacetime depicted in figure 11 and removing a disk of points as shown.

There is another reason why G-prediction seems to mirror only dimly true prediction. We expect that having made a prediction an observer can somehow verify that prediction, either directly by actually visiting that event or indirectly by receiving information from that event at a later time. Geroch implicitly acknowledges this, pointing out that with his prediction — G-prediction — sometimes direct verification is possible and sometimes only indirect verification is possible. For example, only indirect verification of p is possible in the spacetime in figure 11, which can occur when Sibyl reaches an event to future of p, u say.

But in fact a G-predictable event can be absolutely unverifiable. To see this, take the example in figure 11 and remove a three-dimensional disk of points, as depicted in figure 12. Any signal from p will soon vanish into the hole left by the removed disk. Consequently Sibyl can never intersect $J^+(p)$, so the prediction that took place at event q cannot be even indirectly verified.

G-prediction is, in short, too broad, for it can includes cases where the 'predicted' event is not in the future and also cases where verification is absolutely impossible.

The aim now is to formulate a better definition. Suppose we focus first on the case of prediction in which direct verification is possible. Then I suggest that if an event p is 'predictable' from an event q then the following should hold:

(a) q is to the past of p (*pre*diction);

(b) the region that q can causally access contains the data that determines p;

(c) an observer who passes through q must be able to visit p;

(d) p is not in a region q can causally access.

Conditions (a), (b) and (c) are self-explanatory; condition (d) is there to make sure that q cannot acquire knowledge of p directly.

The corresponding mathematical conditions are:

(a′) $q \quad J^-(p)$;

(b′) every past endless causal curve through p must intersect $J^-(q)$;

(c′) $q \quad I^-(p)$;

(d′) $p \quad J^-(q)$.

It is clear that condition (a′) is implied by (c′). Moreover, since strong causality holds, (a′) also implies (d′). Thus (a′), (b′), (c′), and (d′) are equivalent to just (b′) and (c′). These two conditions then are to be satisfied if prediction with direct verification is possible.

On the other hand, if we are concerned with prediction with direct or indirect verification, then condition (c′) would be dropped, leaving just condition (a′), (b′), and (d′). These three are seen to be equivalent to just (a′) and (b′). Comparing this with the previous paragraph reveals that indirectly verifiable events that are not directly verifiable must lie on (not in) q's future null cone; see figure 13.

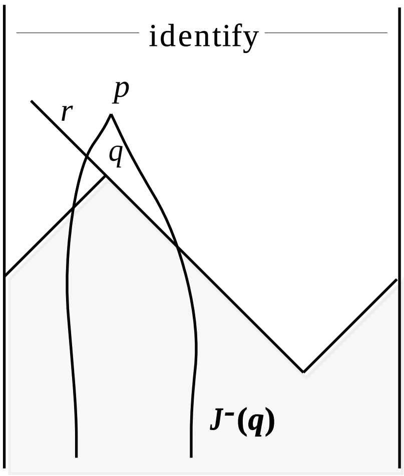


Figure 13. A spacetime formed by identifying the edges of a spatially finite segment of 2-dimensional Minkowski spacetime. p is predictable from q and direct verification is possible (conditions (b′) and (c′)); r however is predictable from q but only indirect verification is possible (conditions (a′) and (b′)).

Since the difference between the sets of directly verifiable and indirectly/directly verifiable predictable events is slight and of no great theoretical importance, I will assume hereafter that prediction entails direct verifiability. This suggests the following definition.

Definition 2.2.1. In a strongly causal spacetime (M, g_{ab}), define the *domain of prediction* of a point q, $P(q)$, as the set of points p such that

(i) every past endless causal curve through p intersects $J^-(q)$,

(ii) $q \quad I^-(p)$.

Moreover, if the spacetime (M, g_{ab}) possesses points p and q such that $p \quad P(q)$, then I shall say that (M, g_{ab}) is a *locally predictable* (LP) spacetime, and p is a *locally predictable* (LP) event.

In the spacetime pictured in figure 14, the event p is predictable from the event q. But not every event is predictable. Event z, for example, is not predictable because the past endless null geodesic μ fails to intersect $I^-(z)$, and *a fortiori* fails to intersect

$J^-(x)$ for any x $I^-(z)$. Thus there is no x for which $P(x)=z$. Moreover, it is also true that $P(z)=$, which means that no predictions whatsoever can be made at z.

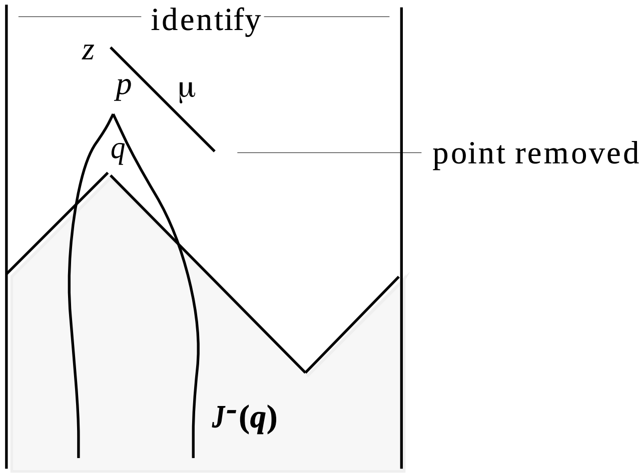
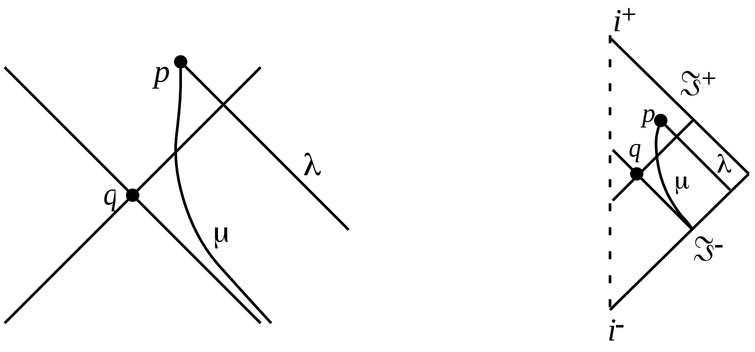


Figure 14. A LP spacetime formed by identifying the edges of a spatial segment of 2-dimensional Minkowski spacetime, and removing a point. p is predictable from q, i.e. p $P(q)$; but z is not predictable from any event. Moreover, $P(z)=$.

That $P(q)$ is empty in Minkowski spacetime follows from the fact that given any two points p and q in this spacetime, either q $I^-(p)$ or there exists a past endless causal curve through p which fails to intersect $J^-(q)$. Two such curves are shown in the usual representation of Minkowski spacetime depicted in figure 15(a). One can represent how these curves do not enter $J^-(q)$ somewhere 'off the page' by using the *Penrose diagram* or, as it is now known the *Carter-Penrose diagram* (Hawking and Penrose 1995, p. 41), of Minkowski spacetime, depicted in figure 15(b). In this representation infinity is brought to a finite distance by means of a conformal transformation that preserves causal relations (Hawking and Ellis 1973), pp. 120-123). The severity of the reduction allows one to adjoin points at infinity to the diagrams. So $^+$ represents the terminus of outgoing null lines; i^+ the terminus of timelike geodesics. (These points do not correspond to anything in spacetime or, *a fortiori*, to anything in the world; they are mathematical constructs only.) Since the LP conditions are cauched in purely causal terms, we have the following: A *spacetime* (M, g_{ab}) *is LP if and only if the 'Carter-Penrose diagram of* (M, g_{ab})*' is LP*

The scare quotes are there to remind us that a Carter-Penrose diagram is, strictly speaking, not a spacetime — so it cannot, strictly speaking, be LP.



(a) (b)

Figure 15. In Minkowski spacetime (a), λ and μ each go to past infinity and never intersect $J^-(q)$. This is confirmed in the corresponding Carter-Penrose diagram (b).

2.3 The structure of locally-predictable spacetimes

There is another way to view the domain of prediction based on the existence of a specific achronal set that determines the predictable event. Suppose that in a spacetime (M, g_{ab}) there are points p, q M such that p $P(q)$. Then every past endless causal curve through p enters $J^-(q)$, but since p $J^-(q)$ (for otherwise there would be causality violations), each such curve must intersect the boundary of $J^-(q)$, $\partial J^-(q)$. Let $S=\partial J^-(q)$. Now since S is an achronal set (Hawking and Ellis 1973, proposition 6.3.1), we see that LP condition (i) is equivalent to requiring that p $D^+(S)$. Coupling this to LP condition (ii) gives the following:

Lemma 2.3.1. $P(q)$ $D^+(S) \cap I^+(q)$.

I have three remarks concerning this result.

First, I do not know whether the ‘ ’ can be replaced by ‘=’.[10] It is certainly conceivable that $D^+(S) \cap I^+(q)$ could be larger than $P(q)$, because there could be a point y in $I^+(q)$ such that a past endless causal curve through y intersects S but not $J^-(q)$ (remember that in general $S=\partial J^-(q) \not\subset J^-(q)$). But I cannot find an example of this.

Secondly, since in general $\partial J^-(q)$ $J^-(q)$, not all the determining information on $\partial J^-(q)$ may be accessible by q. (This occurs in the spacetime depicted in figure 16.)

---

[10] I unjustifiably asserted equality in Hogarth (1993).

Finally, it would be nice if '$P(q)\ \ D^+(\Sigma)\cap I^+(q)$', where $\Sigma=E^-(q)$ because in that case the determining data could be received by q on light rays. However, in general this is false. (Again, the spacetime depicted in figure 16 provides a counter-example.)

Showing that (*G*2) *is false*

Recall that (G2) is Geroch's claim without proof that every LP spacetime is closed in the sense that it must possess a compact slice. The following counter-example reveals that it is false. (After several unsuccessful attempts at proving this claim, I turned to Geroch himself. In the course of our discussions, he discovered a counter-example, which I have adapted into the one given here.)

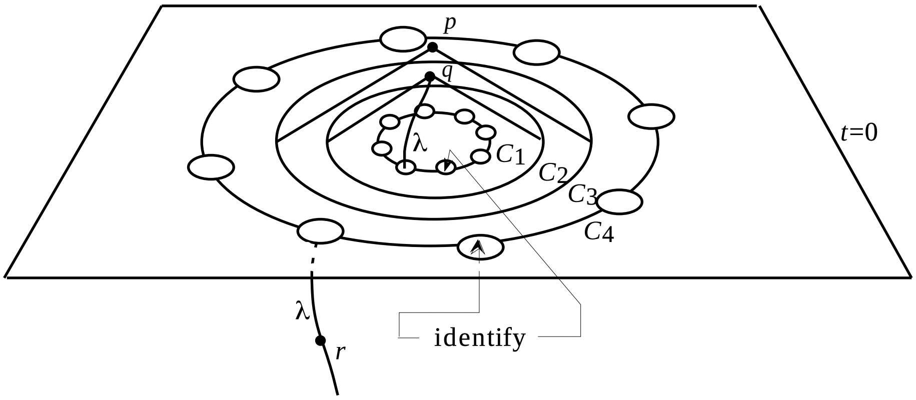


Figure 16. A LP spacetime with no compact slice. Due to an effective 'widening' of q's past light cone, the point $r\ \ J^-(q)$.

To create the counter-example, we start with 3-dimensional Minkowski spacetime, and draw four concentric circles C_1, C_2, C_3, and C_4 in the plane t=0, as depicted in figure 16 (which is not drawn to scale). We can take the circles' respective radii to be 1, 2, 3 and 4. The point p (respectively, q) is given by the apex of the past light cone whose intersection with t=0 is C_3 (respectively, C_2). Now remove 8 equally spaced disks of radii 0.01 on C_1 that lie within the plane t=0, and do the same for eight disks of radii 0.01 on C_4. Next, identify the upper face of the spacetime (where it used to meet a disk) for each disk on the C_1 with the lower face of the spacetime for the corresponding disk on C_4. (To keep the figure simple, I have indicated only one identification, but of course there are eight in all.) The rims of all 8 identified disks are then removed in the manner of figure 11.

I now claim that every past endless causal curve from p enters $J^-(q)$. Intuitively this is so because the identifications have produced an effective 'widening' of q's past light cone to such an extent that eventually it envelopes the past light cone of p. This becomes evident if we examine, say, the plane t=–5 (labelled $\Sigma_{t=-5}$, pictured in figure

17, and again not drawn to scale), by which time $J^-(p)$ is completely contained in $J^-(q)$.

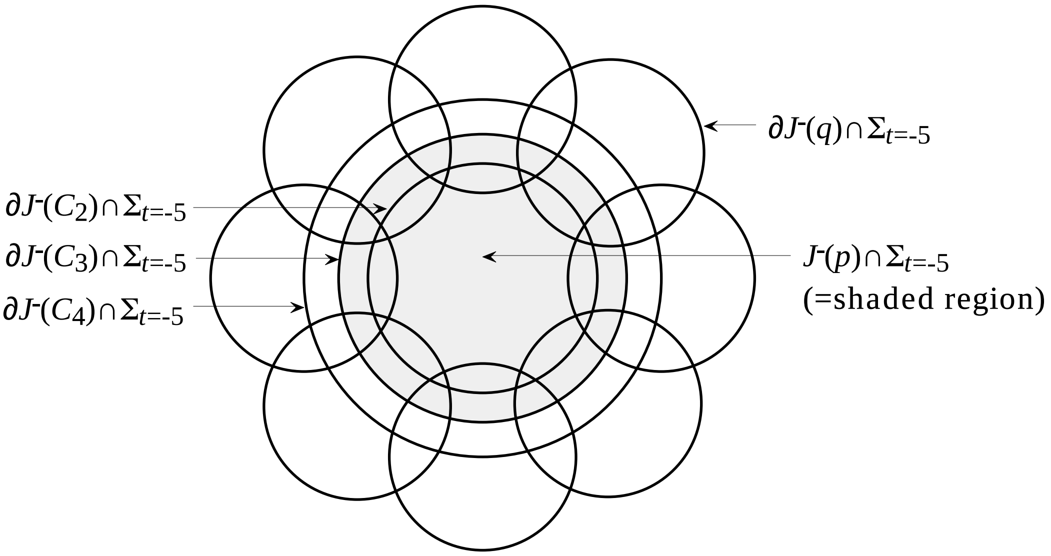


Figure 17. The set $J^-(q)\cap\Sigma_{t=-5}$ is the total 'flower'. It contains the set $J^-(p)\cap\Sigma_{t=-5}$, represented by the shaded region.

This spacetime is therefore LP, but, of course, it has no compact slice (or anything remotely resembling that). We note also that it is stably causal (implies strongly causal), since all identifications take place on the plane t=0.

We can rescue something of Geroch's claim by imposing an additional condition on the LP spacetimes. We first recall that for a closed achronal set S in a spacetime (M, g_{ab}), the *future Cauchy horizon* of S, is labelled $H^+(S)$ and defined by

$$H^+(S) = clos\,[D^+(S)] - I^-[(D^+(S)].$$

The *Cauchy horizon* of S, $H(S)$, is the union of $H^-(S)$ and $H^+(S)$. It is a standard result that $\partial D(S)=H(S)$ (Wald 1984, proposition 8.3.6). Recall also that for a set R in (M, g_{ab}), the set $J^+(R)–I^+(R)$, called the *future horismos* of R, is labelled $E^+(q)$. Finally, I note a simple identity: $\partial I^-(q)=\partial J^-(q)$.

Proposition 2.3.2 (Hogarth 1993). Suppose (M, g_{ab}) is an LP spacetime with points p, q such that p $P(q)$. Suppose further that $\partial J^-(q)=E^-(q)$. Then (M, g_{ab}) admits a compact slice.

Proof. Let $S=\partial J^-(q)=E^-(q)$. By proposition 6.3.1 in Hawking and Ellis 1973, S is a slice. It will be shown that S is compact. By lemma 2.3.1 p $P(q)$ implies p $D^+(S)\cap I^+(q)$ $D(S)$. I claim that p $int(D(S))$. Suppose for a *reductio* that p $int(D(S))$. Then p $\partial D(S)=H(S)=H^-(S)$ $H^+(S)$. But p $I^+(S)$ implies p $I^+(D^+(S))$

implies p $H^-(S)$. So p $H^+(S)$. Now since S is edgeless (it is a slice), the corollary of proposition 6.5.3 in *ibid.* implies that $H^+(S)$ is a slice (implies achronal) generated by past endless null geodesics. Hence p must lie on one such past endless null geodesic that fails to meet $I^-(p)$ and *a fortiori* fails to meet $J^-(q)$ $I^-(p)$ — which contradicts the hypothesis that p $P(q)$. Hence p $int(D(S))$. In this case, proposition 6.6.6 in *ibid.* implies that $J^+(S) \cap J^-(p)$ is compact. This compact set contains S since, obviously S $J^+(S)$, and $S=E^-(q)$ and q $I^-(p)$ implies that S $J^-(p)$. S is therefore a closed set (it is a slice) contained in a compact region, which means S itself must be compact.♦

Although the condition of $\partial J^-(q)=E^-(q)$ can be violated in some spacetimes (e.g. the counter-example above, Minkowski spacetime with a point removed, anti-de Sitter space and the Reissner-Nordström solution), it is trivially satisfied in so–called *causally simple* spacetimes (*ibid.*, p. 188) — spacetimes for which $\partial J^-(p)=E^-(p)$ and $\partial J^+(p)=E^+(p)$ at every point p. It is a standard result that all globally hyperbolic spacetimes (that is, spacetimes that admit a Cauchy surface) are causally simple (*ibid.*, proposition 6.6.1). We saw in chapter 1 that strong cosmic censorship is equivalent to global hyperbolicity, so we have the following:

Corollary 1 (Hogarth 1993). If (M, g_{ab}) is a LP spacetime which satisfies strong cosmic censorship, then (M, g_{ab}) is closed in the sense that (M, g_{ab}) admits a compact slice.

Corollary 2. Minkowski spacetime, the k=0,–1 Robertson-Walker spaces and the Schwarzschild solution are not LP spacetimes.

Proof. Each of these solutions is globally hyperbolic but does not possess a compact slice.♦

Showing that (*G*3) *is false*

The relevant statement is given by Geroch (1977, p. 91): '[point x is G-predictable from q] if and only if $I^-(x) \not\subset I^-(q)$, and, in addition, there is a three-dimensional, achronal surface S in $I^-(q)$ with x in $D^+(S)$.' (G3) is the 'only if' part of this statement. The 'if' part is clearly true, but although (G3) rings true it is actually false.

Consider the following: 'x is in $P(q)$ only if there is a three dimensional, achronal surface S $J^-(q)$ with x in $D^+(S)$'.

It would be doubly interesting for us if this statement turned out to be false. For on the one hand it would disprove (G3) (because (a) p $P(q)$ implies p is G-predictable from q, and (b) S $I^-(q)$ implies S $J^-(q)$; the falsity of (G3) would then follow by *modus tollens*); and on the other, it would say something interesting about LP spacetimes. Of course it is false.

To create the counter-example, take the spacetime depicted in figure 16 and remove two points in the manner shown in figure 18. Since every past endless causal curve through p that reaches either hole must intersect $J^{-}(q)$, it follows that the removal of these two points preserves p $P(q)$.

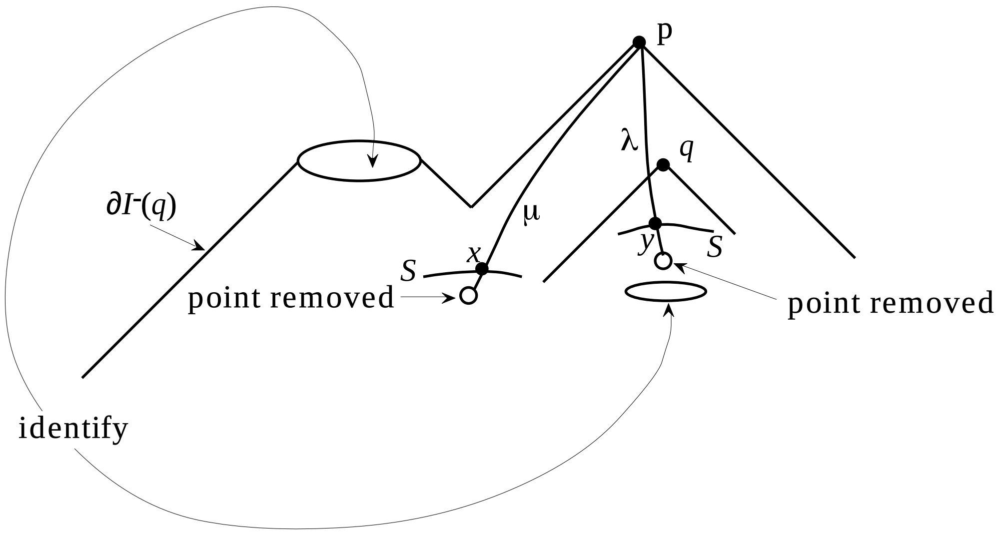


Figure 18. A ‘detail’ of the spacetime depicted in figure 16 but with two extra points removed.

Now suppose there is an achronal set S $J^{-}(q)$ such that p $D^{+}(S)$. Then in order for S to catch the causal curve λ, S must intersect λ before the hole on the right is reached, at some point y say, as depicted in figure 18. Similarly S must also intersect the causal curve μ before the hole on the left is reached, at some point x say. But if the hole on the right is close to q (as we can choose it to be), then any such y can be joined to any such x by a timelike curve that passes through the identified disks. Hence S is not achronal.

(The achronal set given by $\partial I^{-}(q)$ determines p, in accordance with lemma 2.3.1, but it does not lie in $J^{-}(q)$.)

In addition to disproving (G3), this example reveals that: *if a prediction is possible then the relevant data cannot necessarily be read off a single past achronal surface.* Thus in some universes prediction must entail consulting a range of achronal surfaces or perhaps even a spatio-temporal ‘block’.

2.4 Future-set predictable spacetimes

Since in principle we can affect and reach all future events, it is natural to wonder what kind of spacetimes allow at least one observer to predict her entire future. This is, I presume, the motivation that lead Müller-Hoissen (1981) to consider spacetimes (M, g_{ab}) that admit at least one event q with the following property: each past endless causal curve that intersects $I^+(q)$ also intersects $I^-(q)$. I shall call such events, *future-set predictable* (FSP) events; spacetimes that possess at least one FSP event are then *FSP spacetimes*. An example of an FSP spacetime is depicted in figure 19.

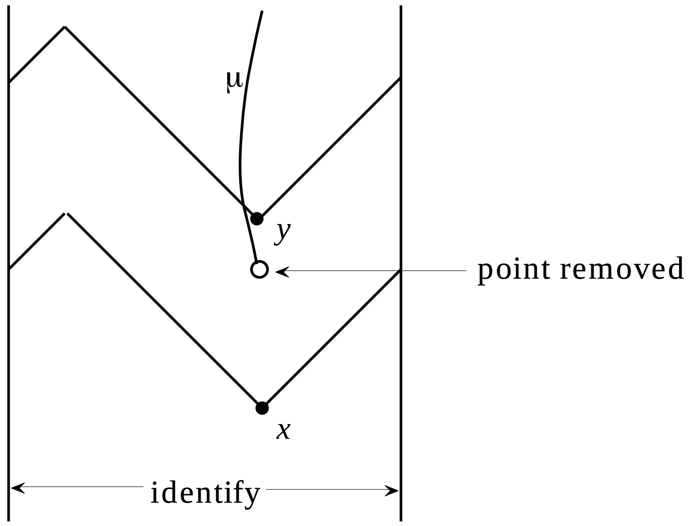


Figure 19. A FSP spacetime obtained by identifying the edges of a spatial segment of 2-dimensional Minkowski spacetime and removing a point. y is an FSP event. x is not an FSP event since the past endless causal curve μ intersects $I^+(x)$ but fails to intersect $I^-(x)$.

As I just hinted, the term ‘future-set’ refer to the fact that an observer at an FSP event can predict not just *some* future events (as can happen in LP spacetimes) but *all* events in her chronological future. Here we understand ‘predict’ in the exact sense of LP spacetimes. That is: if q is an FSP event, then $I^+(q)$ $P(q)$. (This would also have followed if we had defined an event q to be FSP if each past endless causal curve that intersects $I^+(q)$ also intersects J^-(q). In the light of remarks above, this definition is obviously better motivated than Müller-Hoissen’s. However, I shall follow Müller-Hoissen since he is consistent with another paper I want to discuss, viz. Budic and Sachs 1976.)

Clearly every FSP spacetime is an LP spacetime; however the spacetime depicted in figure 20 shows the converse is false.

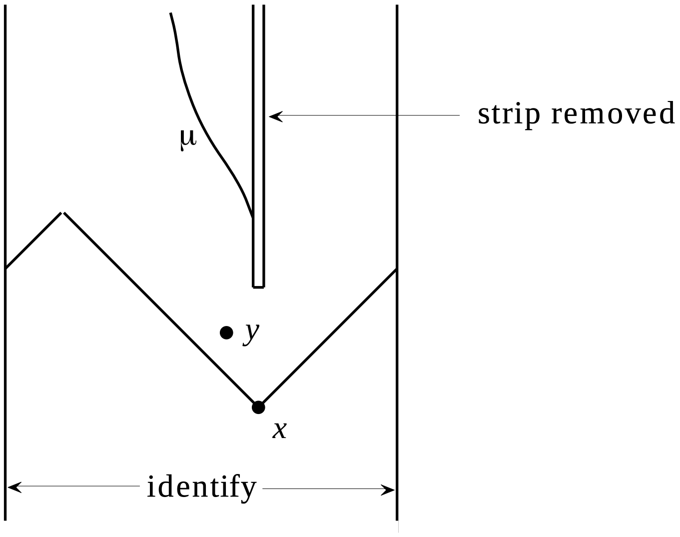


Figure 20. A spacetime formed by identifying the edges of a spatial segment of 2-dimensional Minkowski spacetime and removing an infinite strip. The spacetime is LP since y $P(x)$; but it is not FSP since for any point, e.g. x, there is always a past endless causal curve, e.g. μ, that intersects $I^+(x)$ but does not intersect $I^-(x)$.

In his paper Müller-Hoissen claims to prove that every FSP spacetime admits a compact slice. But like Geroch's similar claim about LP spacetimes, it too is false.[11] Curiously Müller-Hoissen seems to have made the mistake without knowing of Geroch's 'result'; at least he makes no reference to him. Cosmic resonance, it seems, relays false ideas too. The counter-example appears in figure 21.

This spacetime is constructed by taking the spacetime pictured in figure 16 and removing a spacelike disk of points, which lies just to the future of q and which eclipses $J^+(q)$. The point q is FSP because any past endless causal curve that intersects $I^+(q)$ cannot avoid (eventually) intersecting the artificially widened $I^-(q)$.

---

[11] Those who seek the error in Müller-Hoissen's proof should note that his lemma 1.3 cannot be applied to his lemma 1.2.

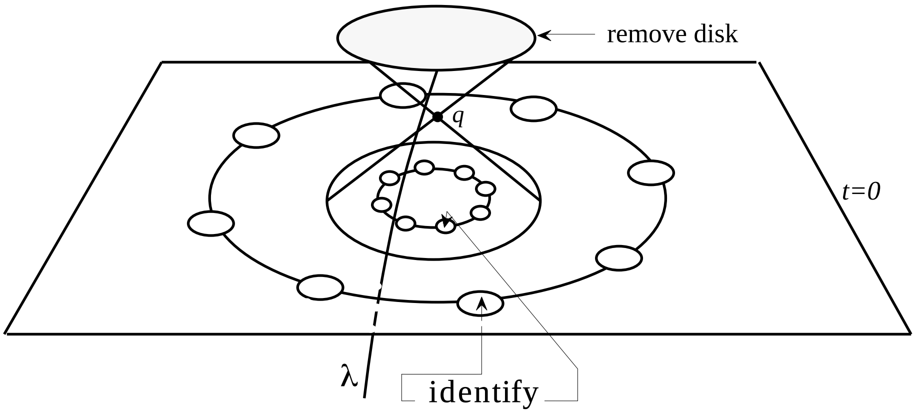


Figure 21. A FSP spacetime with no compact slice.

Again, a result of sorts can be salvaged, for we can couple proposition 2.3.2 with the fact that every FSP event requires at least one LP event to give the following:

Proposition 2.4.1 (Hogarth 1993). Suppose (M, g_{ab}) is an FSP spacetime with FSP event q. Suppose further that $\partial J^-(q)=E^-(q)$. Then (M, g_{ab}) admits a compact slice.

Corollary. If (M, g_{ab}) satisfies strong cosmic censorship and is FSP, then (M, g_{ab}) is closed in the sense that it admits a compact slice.

2.5 Generalised predictable spacetimes

It is natural to wonder what kind of spacetimes are such that every event is LP — what I shall call *everywhere locally predictable* (ELP) spacetimes; or such that every event is FSP — what I shall call *everywhere future set predictable* (EFSP) spacetimes. In ELP spacetimes every event can be predicted from some past event; while in EFSP spacetimes every event can predict its own future.

In this section I will first relate EFSP spacetimes to the notion of a TIF. This will lead to the equivalence of the classes of ELP and EFSP spacetimes. I will then consider an apparently stronger condition, called everywhere completely predictable (ECP), and show that it is in fact equivalent to ELP and EFSP.

So the first result we prove concerns EFSP spacetimes and uses the notion of a *terminal indecomposable future-set* (TIF) (I take my definition from Penrose 1979, p. 621; but see also Hawking and Ellis 1973, p. 218). These are points adjoined to the spacetime to represent either ‘past singularities’ or ‘points at past infinity’. (An analogous definition exists for a *terminal indecomposable past-set*.) The idea is to regard two past endless causal curves, μ and ν, say, as terminating at the same point at past infinity if $I^+(\mu)=I^+(\nu)$. These future sets therefore correspond to points at past infinity, and so it is natural to define a TIF as a set of the form $I^+(\lambda)$, where λ is a past endless causal curve. Notice that since every spacetime admits a past endless causal curve, every spacetime possesses at least one TIF. Some spacetimes however possess only one TIF; they can be thought of as having ‘evolved from a single point’. In this case the future of *any* past endless causal curve is the whole spacetime. Contrast this with what happens in, for example, the standard Robertson-Walker big bang models, where the TIFs constitute a spacelike 3-surface (see figure 22), or in Minkowski spacetime, where the TIFs constitute a null 3-surface (see figure 18).

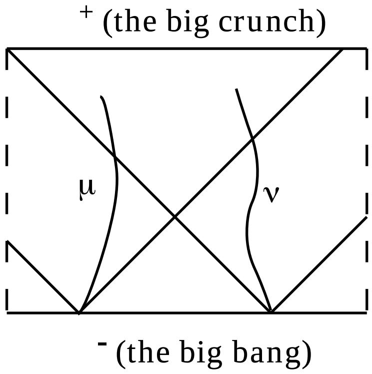


Figure 22. The Carter-Penrose diagram of the closed Robertson-Walker big bang model.

Proposition 2.5.1. A spacetime (M, g_{ab}) is EFSP if and only if (M, g_{ab}) possesses exactly one TIF. (This is stated but unproved by Budic and Sachs 1976.)

Proof. ‘if’ part. Suppose (M, g_{ab}) possesses more than one TIF. Let λ be any past endless timelike curve. Then $M \neq I^+(\lambda)$, by hypothesis. Choose $y \quad \partial I^+(\lambda)$; then theorem 8.1.2 in Wald 1984 implies that y lies on a past endless null geodesic μ contained in $\partial I^+(\lambda)$ (if μ had an endpoint at all it must lie on λ, but that is impossible since λ timelike implies $\lambda \quad I^+(\lambda)$ implies $\lambda \cap \partial I^+(\lambda)= \quad$). Since the set $\partial I^+(\lambda)$ is achronal (Hawking and Ellis 1973, proposition 6.3.1), $\mu \cap I^-(\mu)= \quad$. Now choose x such that $y \quad I^+(x)$. Then μ intersects $I^+(x)$. But μ cannot intersect $I^-(x)$ because $I^-(x) \quad I^-(\mu)$ implies $\mu \cap I^-(x) \quad \mu \cap I^-(\mu)= \quad$. Thus x is not an FSP event, which means (M, g_{ab}) is not a EFSP spacetime.

‘only if’ part. Suppose (M, g_{ab}) possesses exactly one TIF. Choose $x \quad M$, and suppose that some past endless causal curve λ intersects $I^+(x)$. By hypothesis, $M=I^+(\lambda)$, which means that λ must intersect the chronological past of every point in

M. In particular it must intersect $I^-(x)$. So the arbitrarily chosen point x is FSP, and so (M, g_{ab}) is a EFSP spacetime. ♦

From the 'if' part of the proof above we see that not only does the presence of more than one TIF introduce events that are not FSP, but it also introduces events that are not LP. (The point y in the above proof is not LP since μ never intersects $I^-(y)$ and *a fortiori* never enters $J^-(z)$ for any z $I^-(y)$. Thus there is no z for which y $P(z)$.) We conclude that (M, g_{ab}) is ELP only if (M, g_{ab}) possesses exactly one TIF. But the 'if' part of proposition 2.5.1 (and the fact that, trivially, EFSP entails ELP) then ensures:

Corollary A spacetime (M, g_{ab}) is ELP if and only if (M, g_{ab}) is EFSP.

I shall call a spacetime (M, g_{ab}) *everywhere completely predictable* (ECP) if it possesses the property that there exists a Cauchy surface in the causal past of every point. The idea is that in a ECP spacetime the *entire* spacetime is predictable from any event. In this case, one expects that every event can predict its own future; in other words, that every ECP spacetime is a EFSP spacetime. This is our next result.

Proposition 2.5.2 If (M, g_{ab}) is a ECP spacetime, then (M, g_{ab}) is a EFSP spacetime.

Proof: Suppose (M, g_{ab}) is a ECP spacetime; and choose x M. Then for y $I^-(x)$, $J^-(y)$ contains a Cauchy surface, Σ, say. It follows that Σ $I^-(x)$. Now suppose that λ is a past endless causal curve that intersects $I^+(x)$ at z, say. If we extend any causal curve with endpoint z forever into the future, then it cannot meet Σ for that would entail a causality violation. Hence λ must intersect Σ, which implies that λ must intersect $I^-(x)$. Thus the arbitrarily chosen point x is an FSP event, and so (M, g_{ab}) is a EFSP spacetime. ♦

Another important property of ECP spacetimes follows from the next simple lemma. (Geroch (1977, p. 93) remarks that this is due to John Earman, but as far as I am aware a proof has not appeared in the literature.)

Lemma 2.5.3 (Earman). Suppose (M, g_{ab}) is a globally hyperbolic spacetime with Cauchy surface Σ. Suppose further that there is a point x M such that Σ $J^-(x)$. Then Σ is compact.

Proof: Since (M, g_{ab}) is globally hyperbolic, $M{=}D(\Sigma){=}int(D(\Sigma))$ (since M is open). Proposition 6.6.6 in Hawking and Ellis 1973 implies that $J^-(x) \cap J^+(\Sigma)$ is either compact or empty. But this set contains Σ since Σ $\Sigma \cap J^+(\Sigma)$ $J^-(x) \cap J^+(\Sigma)$. Now Σ is closed by definition, so Σ must be compact. ♦

Thus each ECP spacetime (M, g_{ab}) is closed in the strong sense that it must admit a compact Cauchy surface. By a theorem of Geroch (Hawking and Ellis 1973, proposition 6.6.8), a globally hyperbolic spacetime (M, g_{ab}) with Cauchy surface Σ must have topology $R\times\Sigma$. So every ECP spacetime is foliated by topologically equivalent compact surfaces. We note, however, that not every spacetime that admits a compact Cauchy surface is ECP. Consider the spacetime formed by identifying the edges of a spatial segment of Minkowski spacetime. Now delete all points for which $t\leq 4004$ BC. This truncated spacetime, depicted in figure 23, admits a compact Cauchy surface, but events near the moment of 'creation' do not possess a Cauchy surface in their causal pasts.

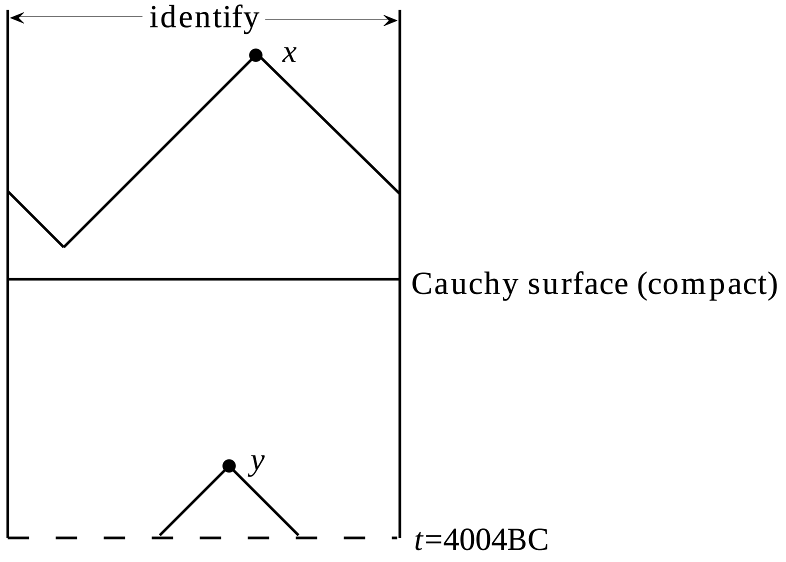


Figure 23. Truncated cylindrical spacetime. It possesses a compact Cauchy surface but is not ECP because although some events (e.g. x) have a Cauchy surface in their past, some others do not (e.g. y).

The next proposition reveals something of the nature of the Cauchy surfaces in EFSP spacetimes. (In what follows, I make use of the identity $int(D(S))=I^-(D^+(S))\cap I^+(D^-(S))$, which can be found in Wald 1984, p. 202.)

Proposition 2.5.4. Suppose (M, g_{ab}) is a EFSP spacetime. Then for y M, $E^-(y)$ is a compact Cauchy surface.

Proof: We first show that (M, g_{ab}) is globally hyperbolic. Let p and q be any two points such that $J^-(p)\cap J^+(q)$ is non-empty, and choose y $I^-(p)$. Let $S=\partial I^-(y)$. Since, by hypothesis, y is FSP, every past endless causal curve that intersects $I^+(y)$ must intersect $I^-(y)$ and *ipso facto* must intersect S; in other words, $I^+(y)$ $D^+(S)$. Now by proposition 6.6.3 in Hawking and Ellis 1973, $int(D(S))$ is globally hyperbolic, and since $int(D(S))=I^-(D^+(S))\cap I^+(D^-(S))$ $I^-(I^+(y))\cap I^+(S)$ $I^+(y)$, $I^+(y)$ is also globally hyperbolic. Thus $I^+(y)$ is a globally hyperbolic region containing p and q, which implies that $J^-(p)\cap J^+(q)$ is compact. Since this argument runs for any events p and q, (M, g_{ab}) is globally hyperbolic.

By proposition 6.3.1 in *ibid.*, S is a slice. Thus if S is not a Cauchy surface then there must exist at least one endless curve λ such that $\lambda\cap S=$. If this were so, then either (i) $\lambda \quad I^-(y)$ or (ii) $\lambda\cap I^-(y)=$. I claim both cases lead to a contradiction. Suppose that (i) holds, and choose $x \quad \lambda$. The segment of λ lying to the future of x is then totally future imprisoned (*ibid.*, p. 194) in the set $J^+(x)\cap J^-(y)$. Since strong causality holds on (M, g_{ab}), we can apply proposition 6.4.7 in *ibid.* to show that $J^+(x)\cap J^-(y)$ is non-compact, which contradicts the fact that (M, g_{ab}) is globally hyperbolic. Suppose then that (ii) holds. Then $I^+(\lambda)$ and $I^+(\mu)$, where μ is a past endless curve through y, must be distinct TIFs since $I^+(\mu)$ contains y and $I^+(\lambda)$ does not. But this contradicts proposition 2.5.1. We conclude that S must be a Cauchy surface. Finally, we note that since (M, g_{ab}) is globally hyperbolic, proposition 6.6.1 in *ibid.* implies that $S=E^-(y)$, so $S \quad J^-(p)$. Lemma 2.3.3 then implies that S is compact.♦

We see from the above proof that an arbitrarily chosen point p in a EFSP spacetime always contains a Cauchy surface in its causal past. In other words

Corollary If a spacetime (M, g_{ab}) is EFSP, then (M, g_{ab}) is ECP.

Propositions 2.5.1 and 2.5.4 and their corollaries, together with proposition 2.5.2, can be summarised in the following:

Proposition 2.5.5 (Hogarth 1993). The classes of ELP, EFSP, ECP and ‘exactly one TIF’ spacetimes are one and the same.

Most surprising is the equation of the ELP and ECP classes. This reveals that a spacetime which is predictable everywhere in the weakest possible sense that allows direct verification, namely, with each event being predictable from just *some* event in its past, must in fact be completely predictable from *every* event. Put the other way around: *either the whole universe is predictable from any event or there exists at least one absolutely unpredictable event*.

2.6 Examples

Perhaps the simplest example of a ECP spacetime is that obtained by identifying the edges of a spatially finite segment of Minkowski spacetime. Another artificial example (Budic and Sachs 1976, p. 27), is the 2-dimensional cone spacetime depicted in figure 24. In accordance with propositions 2.5.4 and 2.5.5, this spacetime has exactly one TIF and admits a compact Cauchy surface, $\Sigma = E^{-}(x)$. Notice that since Σ' is also a Cauchy surface it must have the same topology as $E^{-}(x)$, viz. S^1

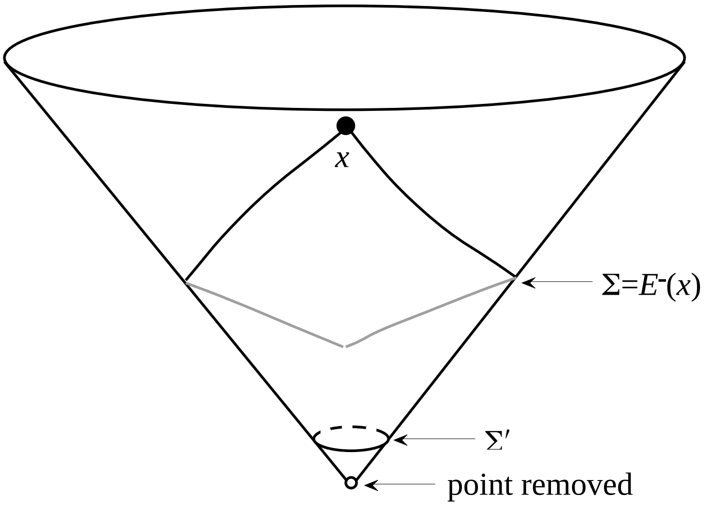


Figure 24. A toy ECP spacetime.

To discover whether there are there are physically reasonable ECP spacetimes, we look to the exact solutions of the EFE, $G_{\mu\nu} = 8\pi T_{\mu\nu}$, where the stress-energy tensor $T_{\mu\nu}$ represents some reasonable form of matter (Hawking and Ellis 1973, chapter 5). In trying to decide whether a given solution (M, g_{ab}) is or is not ECP, we first subject it to the following three-point test. 1. Is $M = R \times \Sigma$, where Σ is compact? 2. Does (M, g_{ab}) admit a compact Cauchy surface? 3. Does (M, g_{ab}) possess exactly one TIF? A negative response to any question means the given solution is not ECP. One can easily check that the following do not make the grade: Minkowski spacetime, Schwarzschild solution, Reissner-Nordström solution, Kerr solution, the standard big bang Robertson-Walker spaces, Taub space, and all asymptotically flat solutions. But ECP spacetimes can in some cases be constructed from these solutions by making appropriate topological identifications. Since the EFE exert only a local constraint, the resulting spacetimes will also satisfy the EFE. One example is the usual cylindrical spacetime, formed by identifying the edges of Minkowski spacetime. Another example is the open Robertson-Walker model with toroidal identifications to make compact spacelike surfaces.[12] It appears though that there are no truly ‘natural’ ECP solutions to the EFE.

---

[12] I am grateful to an anonymous referee of Hogarth (1996) for pointing this out to me.

Exact solutions to the Einstein equations with non-zero cosmological constant are not generally regarded as being physically reasonable. All the same these equations do yield at least two ECP spacetimes: the Einstein static solution and the Eddington-Lemaitre universe. From their Carter-Penrose diagrams (Tipler 1986) pictured in figure 25, it is clear they are ECP because they each possess only one TIF.

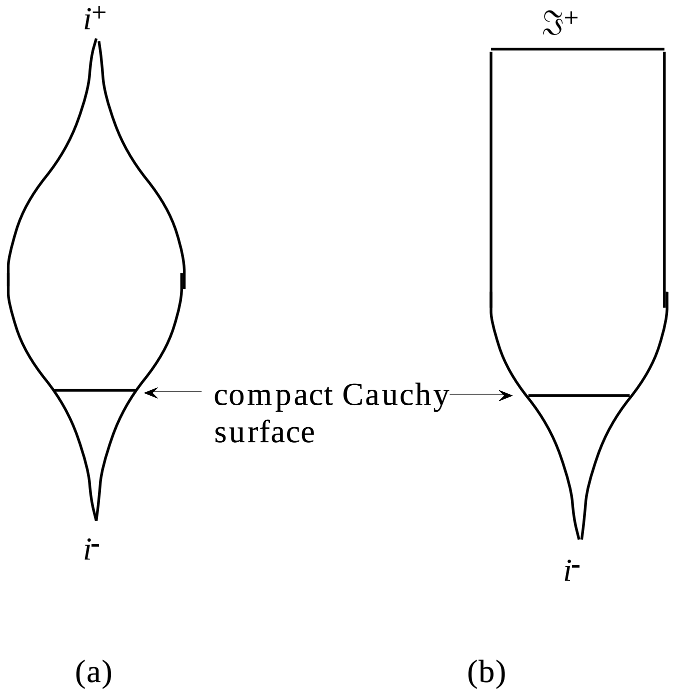


Figure 25. The Carter-Penrose diagram of the Einstein Universe (a), and the Eddington-Lemaitre universe (b).

The Eddington-Lemaitre universe is a closed Robertson-Walker universe that asymptotically approaches the Einstein static universe as $t \to -\infty$, and the de Sitter universe as $t \to \infty$. Consequently, it has the conformal structure of the Einstein universe in the past, and the conformal structure of de Sitter space in the future.

A word about examples of LP spacetimes. We noted in section 2.3 that a number of standard exact solutions are not LP. Other exact solutions that should be checked include the k=+1 Robertson-Walker solution, Taub space, and Kerr metrics. I offer no proof, but an examination of each of these solutions strongly suggests that none are LP. This largely exhausts the important exact solutions, and so it seems that by and large the models of general relativity refuse to accommodate the possibility of prediction.

2.7 World prediction

Previously I have been discussing the possibility of predicting deterministic fields, e.g. the Maxwell field, that were defined *on* spacetime. But in GR, the metrical field — i.e., the spacetime itself — and the matter fields are inextricably woven together (in accordance with the Einstein equations), so one cannot compute the evolution of the matter fields separately from the metrical fields or vice versa. The only hope is to compute the combined evolution of the two kinds of fields. This presents a quite different situation. Previously Sibyl knew at the outset which spacetime point she was to make a prediction about, but now the existence of the points themselves must also be established.[13] I will call the prediction of both the spacetime point itself and its attendant physical state, a *world prediction*.

That world prediction might be possible is suggested by the fact that GR admits a well-posed initial value formulation (see Wald 1984, chapter 10; Geroch and Horowitz 1979, pp. 284-285.) Here is an outline of the key result.

Suppose S a 3-dimensional smooth manifold, and let h_{ab} be a smooth Reimannian metric on S, K_{ab} be a smooth symmetric tensor field on S, and T_{ab} represent various matter fields on S. Intuitively h_{ab} measures the intrinsic spatial geometry of S and K_{ab} describes the way in which S is to be embedded in spacetime. The EFE impose constraints on these h_{ab}, K_{ab}, T_{ab}. Let us suppose these are satisfied[14], in which case S and its attendant tensor fields is called an *initial data set*. The main theorem is this: for any given initial data set, S, there is a spacetime, satisfying the Einstein equations, which has S as a Cauchy surface and such that the data induced on S from the spacetime agrees with the original initial data on S. Moreover, there is a unique maximal spacetime (i.e. a spacetime that cannot be isometrically mapped into a proper subset of another spacetime) with this property.

To ensure that S evolves into this unique maximal spacetime (i.e. not into one of the smaller spacetimes with S as Cauchy surface) we insist that spacetime is *hole-free*, i.e. every spacetime (M, g_{ab}) has the following property: given any achronal set S in M, and any metric-preserving embedding Ψ of $D(S)$ into some other spacetime $(M^*, g^*{}_{ab})$ then $\Psi(D(S))=D(\Psi(S))$; see figure 26.

---

[13] Geroch (1977) and Budic and Sachs' (1976) are both ambiguous about whether they are discussing prediction of the physical fields on a given spacetime or prediction of the physical fields plus spacetime.

[14] Precise technical details are given by Wald (1984, p. 264).

Now call the theory of GR (implies EFE hold) with the constraint that spacetime is hole-free, *hole-free general relativity* (HFGR). Suppose $(M^1,g^1{}_{ab})$, $(M^2,g^2{}_{ab})$ are two models of HFGR, and let S_1 M^1 and S_2 M^2 be two initial data sets. What the two preceding paragraphs show is that if there is mapping Ψ of S_1 and its fields onto S_2, then $\Psi(D(S_1))=D(S_2)$.

In other words, in HFGR an initial data set S will not merely determine what will happen *in* D(S) (this always follows) but what D(S) actually is. In this case I will say that D(S) is *world determined* by S.

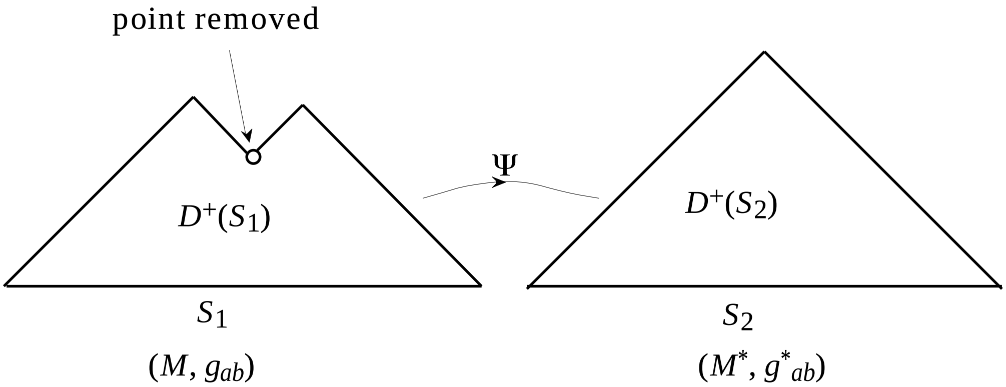


Figure 26. The spacetime on the right, formed by removing a point from Minkowski spacetime, possesses a hole because $D^+(S_1)$ can be isometrically embedded in large spacetime (Minkowski).

The following result provides the link (hinted at in section 2.1) between the previous notion of domain of prediction and the new notion of world prediction.

Result 2.7.1. Within HFGR, the set of spacetimes that allows world prediction is precisely the set of LP spacetimes.

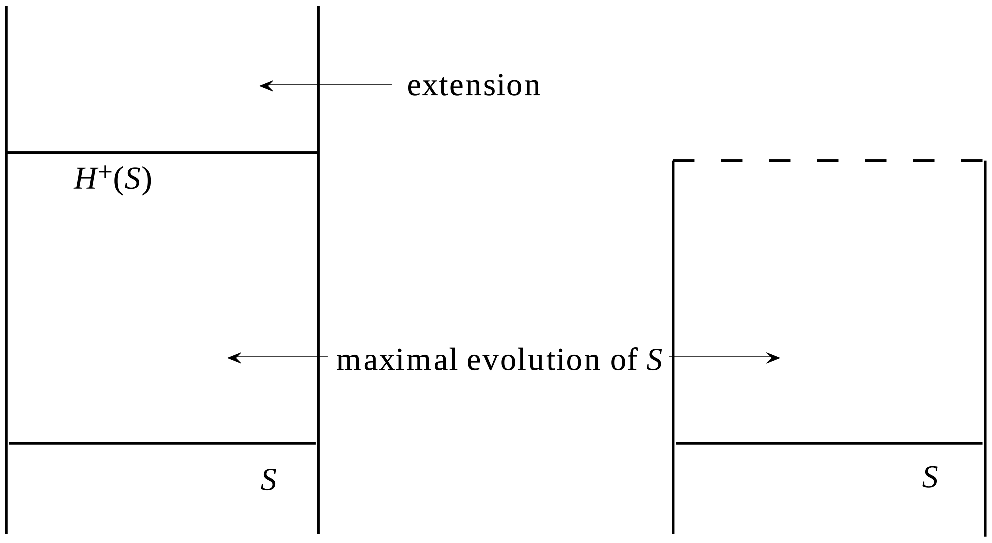
Proof. Assume HFGR. First we will show that world prediction implies LP. So suppose in a spacetime (M, g_{ab}) a world prediction is possible from an event q. Then there must be an achronal set S $J^-(q)$ whose maximal evolution intersects $I^+(q)$. But if p is one such point in the intersection, then it follows that p $I^+(q)$ and every past endless causal curve p intersects S $J^-(q)$. Hence p $P(q)$. In short: if q allows a world prediction of p then p $P(q)$. For the other direction, suppose that in a spacetime (M, g_{ab}) there are points p, q such that p $P(q)$. For a world prediction of p from q, there needs to an achronal set S $J^-(q)$ for which p $D^+(S)$. We have already seen there are exceptional LP spacetimes (figure 18) for which must fail. But by choosing $S=\partial J^-(q)$ the condition p $D(S)$ is satisfied (by lemma 2.3.1) and although in general $S\not\subset J^-(q)$, $S=\partial J^-(q)$ can be *inferred* from the data in $J^-(q)$ by a continuity argument (the fields are assumed continuos remember). This shows that in principle q can make a world prediction of p.♦

Let us return to the hole-free condition. If this is not assumed then world determinism fails and *a fortiori* world prediction fails too. This certainly provides a incentive to adopt the condition, but can it be given some independent justification? It seems not. All one can really do is point to the fact that all the standard solutions of the EFE are hole-free, so HFGR agrees with GR where it matters. For this reason physicists impose hole-freeness by fiat, and I shall follow them.[15]

Imposing the condition that spacetime must be maximally extended (that is, cannot be isometrically embedded into a larger spacetime) also affects prediction. (It might seem as if spacetime maximally-extended implies spacetime hole-free, but that is false. A counter-example is given by Earman 1995, p. 98.) To see how this works, suppose that in a hole-free spacetime (M, g_{ab}) Sibyl has accessed an achronal set S M and established that $D(S)$ is ECP; see figure 27. Then this much is clear: if $D(S)$ is inextendible then Sibyl knows for sure that $M=D(S)$ and so (M, g_{ab}) is ECP. But suppose $D(S)$ is extendible. With the maximally extended condition in place Sibyl knows that (M, g_{ab}) must extended beyond $H^+(S)$, though what the extension is will be beyond her ken. (In general, extensions are not unique. Recall that Taub space, a maximally evolved vacuum solution, has at least two.) If, on the other hand, (M, g_{ab}) is not *a priori* maximally extended, then Sibyl must consider one further possibility: that there is nothing beyond $D(S)$, that is to say $D(S)=M$.

---

[15] The fact that this poses a very real threat to determinism has received little attention in the literature (though see Earman 1995, section 3.8 and Geroch 1977), which is perhaps curious when one reflects on the number of recent journal articles (see Butterfield, Hogarth and Belot 1996 for a bibliography) devoted to shoring up determinism against the possible ruin inflicted by the ‘hole argument’.



(a) (b)

Figure 27. An extendible spacetime with slice S whose maximal evolution is ECP. If spacetime is *a priori* maximally extended, then an extension must exist (a). If not, then another possibility arises: there may be nothing beyond the maximal evolution (b).

To sum up this section: world prediction is possible if and only spacetime is LP and satisfies HFGR. How much of spacetime is predictable depends on the spacetime as a whole and the achronal set chosen. If Sibyl is lucky and computes the evolution of a particular set to be an inextendible ECP spacetime, then nothing worldly will be unseen to her eyes. If, on the other hand, the evolution is an extendible ECP spacetime, then the world could either end where her map ends or could continue across the Cauchy horizon into *terra incognita*. If spacetime is deemed inextendible, then only the latter can hold; if not, then either situation is possible.

2.8 Do predictable spacetimes yield paradox?

I turn now to a more logical issue. Imagine that Sibyl lives in an ECP world and that her calculations reveal that at a future event p the lamp in her bedroom will be switched on. Could she not then ensure that the lamp will be off at p? If yes, then the event was not predictable. If no, then Sibyl is unable to perform an act which a human surely can perform. Either way seems paradoxical. One might wonder if the resolution lies in the vexed concept of free-will, but actually the argument can be given without reference to any sentient being. The trick is to replace Sibyl with an automated robot who is programmed both to predict the future state of the lamp and then to bring about the opposite state of affairs.

The reader may already detect a familiar ring to these words. It is sometimes claimed (see Horwich 1987) that in a spacetime with closed timelike curves (CTCs) an observer or robot or whatever can travel into their own past and paradoxically undo what has been done. The stories are well-known. The man performs autoinfanticide. The bomb destroys its own bomb-making equipment. The robot steps on the chips that will eventually go to make up its brain. The gun shoots itself before the shot is fired. And so on. All the stories have the same form. Start by stating that an event p is such and such (the gun fires the bullet), then follow the trajectory of an object that was at p (the bullet) forward in time round a CTC until it again reaches p. Since p lies in the future, we expect that the object (the bullet) can affect p in all sorts of ways, some of which are contrary to the way p is (destroys the gun).

What these stories — or better, *pseudostories,* because they are genuinely inconsistent and thus represent *nothing* — show is not that CTC worlds are somehow conceptually flawed, but rather that in such worlds the physical fields are massively constrained by the fact that along a CTC the system must evolve back to its original state. There's nothing mysterious about that. Spacetime in general places constraints on its attendant physical fields, so the fact that an ordinary gun cannot find a home in a causally vicious spacetime is no more perplexing than the fact that an elephant cannot find a home in a mouse-sized universe (or a Turing machine in our universe; see section 3.7). (Although the CTC constraints are massive, there are some non-vacuous scenarios consistent with it. Lossev and Novikov (1992), for example, show how a billiard ball can traverse a CTC and collide with its former self in a perfectly consistent way. See also Earman 1995, chapter 6.)

Now the point is that this analysis carries over essentially unchanged to ECP spacetimes. The predictable aspect of these worlds places constraints on the evolution of physical systems which makes contrary lamp switching robots and the like impossible. But this ruling out of some superficially plausible sounding scenarios in no way implies a conceptual flaw. It simply suggests that there are fewer possibilities than usual. (Moreover, since most scenarios are *prima facie* possible in ECP worlds, it is reasonable to suppose that the field constraints involved are weak compared with the mighty constraints of CTC worlds.)

2.9 Prediction and singularities

The affect of singularities on predictability

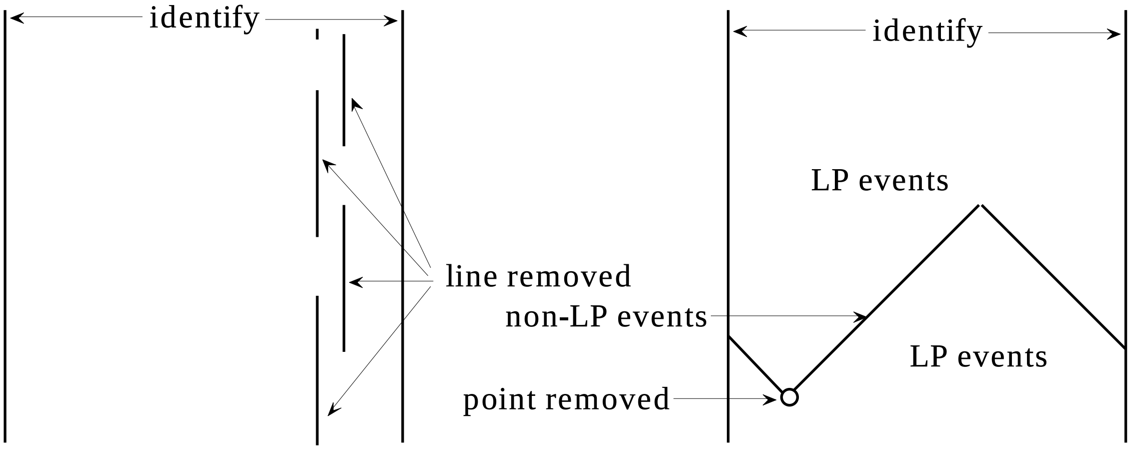
A reader's first brush with the notion of a singularity may leave her with the impression that singularities cause terrible trouble for prediction. Certainly that idea abounds in popular physics books. A randomly chosen example is Davies 1983, p. 56: 'A singularity represents the ultimate unknowable in science. It is an edge or boundary of spacetime at which matter and influences can enter or leave the physical universe in a totally unpredictable fashion'. But clearly this cannot be right because the cone spacetime depicted in figure 24 is ECP and yet has a past singularity So what kind of singularities, if any, do necessarily upset our forecasts?

Proposition 2.9.1. Suppose (M, g_{ab}) is a spacetime which violates strong cosmic censorship. Then there is at least one event p M that is not LP.

Proof. If every event were LP, the (M, g_{ab}) would be ELP. Propositions 2.5.4 and 2.5.5 then show that (M, g_{ab}) must possess a Cauchy surface, which contradicts the assumption that (M, g_{ab}) violates strong cosmic censorship.♦

Corollary Naked singularities necessitate the existence of non-LP events.

Proof. SCC implies no naked singularities. (See corollary of proposition 3.5.2.)♦



(a) (b)

Figure 28. Take one ECP spacetime and add a naked singularity. The naked singularity in (a) spoils prediction throughout, while the naked singularity in (b) spoils prediction on just one 'loop' of events.

What is the extent of the unpredictability inflicted by a naked singularity? The popular idea in one of widespread contamination, but actually it depends on the nature and particularly the size of the singularity involved. Remove an unending series of line segments from the ECP cylindrical spacetime, as in figure 28(a), and the whole universe becomes unpredictable. But remove just a single point (28(b)) and the set of non-LP events is but a single 'loop'; the rest of the universe remains perfectly predictable.

The affect of prediction on singularities

While these results are roughly in line with our intuitions, proposition 2.9.3 (coming up shortly) is not something one would immediately guess. It states, loosely speaking, that universes (by which is meant 'spacetimes that satisfy some very weak physical conditions') that possess at least one LP event (the minimal sense in which a universe is deemed predictable) must be geodesically incomplete, i.e. possess a singularity. It's ironic really. The popular view is that singularities oppose prediction, but in a sense the opposite is true: the possibility of prediction *necessitates* the existence of singularities.

I will outline the statement of this result, and then prove it formally in the next section.

Informal rendition of a *singularity theorem for LP spacetimes*

In chapter 1 I sketched the proof of a simple but weak singularity theorem. Now I need recourse to considerably a more powerful version due to Hawking and Penrose (1970). Informally this can be stated like this.

A spacetime (M, g_{ab}) which satisfies the following cannot be timelike or null geodesically complete:

(a) there are no closed timelike curves;

(b) close-by timelike and null geodesics have a tendency to draw closer;

(c) every geodesic encounters some curvature, however small;

(d) there is either a compact slice or a point at which null geodesics are eventually focused.

These geometrical conditions can be given the following physical interpretation. Condition (a) says that no observer can travel into her own past (recall the alleged paradox involving retro-autoinfanticide). Condition (b) says that gravity is in some weak sense ‘attractive’ (this will hold if, for example, the EFE and the strong energy condition hold). Condition (c) fails to hold only in very ‘special’, highly symmetric spacetimes. Finally, in condition (d), the compact slice is a feature of any spatially closed universe, and the focused null rays are believed to exist in regions where a sufficiently massive body (e.g. a star) is undergoing gravitation collapse (it is basically the definition of a black hole). Conditions (a), (b), and (c) are generally thought to hold in any physically reasonable spacetime, but whether condition (d) holds needs to be checked from model to model.

Returning to LP spacetimes, I now ask: is every LP spacetime singular, i.e. geodesically incomplete? The answer is clearly no for the cylindrical spacetime depicted in figure 13 provides an obvious counter-example. But suppose attention is restricted to LP spacetimes that satisfy the above weak physical conditions (a), (b), and (c). Then the answer is yes. And that is the main result (=proposition 2.9.3): every LP spacetime that satisfies conditions (a), (b), and (c) is singular.[16] Since any model of the universe is expected to satisfy these three conditions, the result can be expressed thus: *universes with at least one predictable event must possess* a *singularity*. Consequently to predict just one event just one second in the future requires that spacetime must be torn off short!

Now proposition 2.9.3 proves that LP universes are singular, but another result (corollary 2 of lemma 2.9.2 below) actually shows that in any LP spacetime there is a well-defined region ‘between’ the event at which the prediction occurs and the event being predicted which must be non-singular (non-singular in a sense to made precise). I call this the *anticipation region* (again, to be made precise), so-called because an observer in this region is either already anticipating, or can manoeuvre into a position and begin to anticipate, what will happen at the predicted event. See figure 29. Proposition 2.9.3 and corollary 2 combine to show that in LP universes the complement of the anticipation region must be singular. This, incidentally, does not imply the existence of unpredictable events in the complement of the anticipation region. That would necessarily follow only if the expected singularity was ‘naked’, and there is nothing to suggest it is. In fact I have no information about this singularity; and in particular, I do not know whether it resides in the past (like the ‘big-bang’) or in the future (like the ‘big-crunch’).

---

[16] The spacetime in figure 13 violates condition (c).

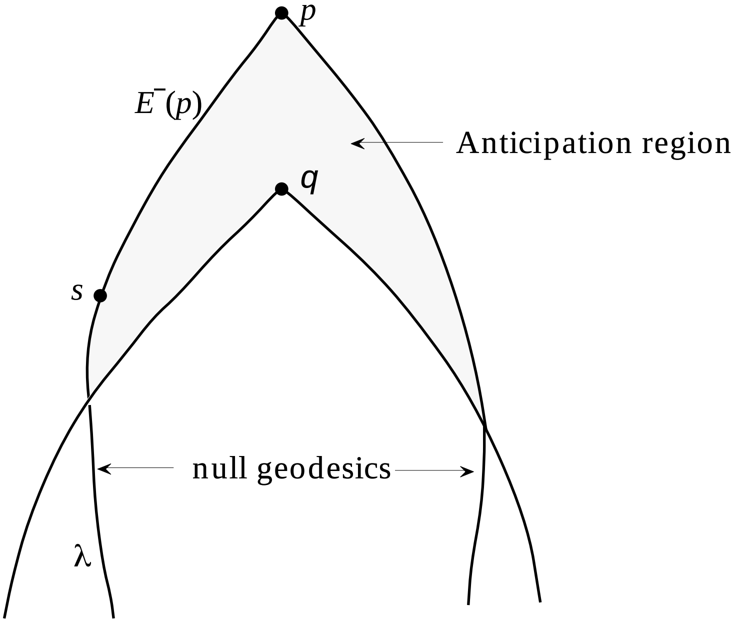


Figure 29. An LP spacetime in which p is predictable from q. Since influences on p must be registered in q's past, p's past light cone must eventually enter q's past light cone. The anticipation region is roughly the set of points in p's past that are not also in q's past. The region cannot possess any singularities. This diagram is used in the proof of lemma 2.9.5.

To summarise: if an event is predictable then the corresponding anticipation region must be non-singular and the complement of that region must be singular. The intuitive idea that singularities oppose prediction is therefore true with respect to the anticipation region, but ironically it is essentially the negation of that idea that holds for the complement.

All that remains now is to prove formally the results described above. For the sake of an easy proof of proposition 2.9.3, it would be nice if every LP spacetime obviously satisfied condition (d), that is, that there is a compact slice or null ray focusing; but this is not so. (Geroch's (1977) claim that every LP spacetime must admit a compact slice would have done the trick, but, as we have seen, this is false.)

Formal rendition of a singularity theorem for LP spacetimes

Hitherto I have assumed for the sake of simplicity that strong causality holds, but now I want to allow the possibility of causal violations. The reason for this will shortly become clear. $P(q)$ is defined as before i.e. as the set of event p such that

(i) every past endless causal curve through p intersect $J^-(q)$,

(ii) q $I^-(p)$,

(iii)p $J^-(q)$,

but notice that now (iii) is not implied by (ii).

In a LP spacetime (M,g_{ab}) with points p, q M such that p $P(q)$, let $\Sigma=\partial J^-(q)$ $(=\partial I^-(q))$, and define the *Anticipation Region* of p and q, denoted $AR(p,q)$, as the set $I^+(\Sigma)\cap I^-(p)$. See figure 29.

(In what follows, I will assume the simple identities $clos\{I^-(x)\}=clos\{J^-(x)\}$, $\partial J^-(x)=\partial I^-(x)$, $int\{I^-(x)\}=I^-(x)$, $int\{D^+(S)\}=I^-(D^+(S))\cap I^+(S)$, $E^-(x)\cap I^-(x)=$, and $\partial I^-(x)\cap I^-(x)=$.)

Lemma 2.9.5. Suppose a spacetime (M, g_{ab}) is LP with points p, q such that p $P(q)$. Then $AR(p,q)$ is globally hyperbolic.

Proof. If $\Sigma=\partial J^-(q)=clos\{J^-(q)\}-int\{J^-(q)\}$ were empty then $J^-(q)$ would be both open and closed, so either $J^-(q)=M$ or $J^-(q)=$. But the first of these cannot hold because by condition (iii) p $J^-(q)$; and the second cannot hold because q $J^-(q)$. Thus Σ is non-empty.

Now let λ be a past endless causal curve through p, and suppose that λ failed to intersect Σ. In that case, we have $\lambda\cap\Sigma=$ implies $\lambda\cap[clos\{I^-(q)\}-I^-(q)]=$ implies $\lambda\cap clos\{I^-(q)\}\cap[M-I^-(q)]=$. This means that $\lambda\cap clos\{I^-(q)\}=$ or $\lambda\cap(M-I^-(q))=$ or $clos\{I^-(q)\}\cap[M-I^-(q)]=$. But none of these three statements can hold because the first contradicts condition (i), the second implies λ $I^-(q)$ which contradicts condition (iii), and the third contradicts the conclusion of the previous paragraph, that Σ is non-empty. Thus every past endless causal curve through p intersects Σ.

Hence from the definition of $D^+(\Sigma)$ and the fact that Σ is an achronal set (Hawking and Ellis 1973, proposition 6.3.1), it follows that p $D^+(\Sigma)$. But $int\{D^+(\Sigma)\}=I^-(D^+(\Sigma))\cap I^+(\Sigma)$ $I^-(p)\cap I^+(\Sigma)$, and according to proposition 6.6.3 in *ibid.*, $int\{D^+(\Sigma)\}$ is globally hyperbolic, so $I^-(p)\cap I^+(\Sigma)=AR(p,q)$ is globally hyperbolic too.♦

Corollary 1. Any anticipation region is strongly causal.

Proof. A globally hyperbolic region is necessarily strongly causal (*ibid.*, lemma 6.6.4). ♦

Corollary 2. Given any two points x and y in $AR(p,q)$, $J^+(x)\cap J^-(y)$ cannot contain a future endless or past endless causal curve.

Proof. Since $AR(p,q)$ is globally hyperbolic, $J^+(x)\cap J^-(y)$ is compact or empty. But by corollary 1, strong causality holds on $AR(p,q)$, so we can apply proposition 6.4.7 in *ibid.* to show that $AR(p,q)$ cannot contain a causal curve that is future endless or past endless. ♦

The respective physical interpretations of corollaries 1 and 2 are that AR possesses no causality violations and no singularities. This does not rule out singularities in the spacetime as a whole, and indeed our next result shows that in any physically reasonable LP spacetime singularities are expected.

Proposition 2.9.3. A spacetime (M, g_{ab}) is not timelike and null geodesically complete if:

(1) there are no closed timelike curves in (M, g_{ab});

(2) $R_{ab}K^aK^b \geq 0$ for every non-spacelike vector K^a;

(3) the generic condition is satisfied, i.e. every non-spacelike geodesic with tangent vector K^a contains a point at which $K_{[a}R_{b]cd[e}K_{f]}K^cK^d \neq 0$;

(4) (M, g_{ab}) is LP.

(Conditions (1), (2) and (3) correspond, respectively, to conditions (a), (b) and (c) in the informal rendition above.)

The proof follows immediately from the next two results.

Proposition 2.9.4 (Hawking and Penrose 1970). A spacetime (M, g_{ab}) is not timelike and null geodesically complete if:

(1) there are no closed timelike curves in (M, g_{ab});

(2) $R_{ab}K^aK^b \geq 0$ for every non-spacelike vector K^a;

(3) the generic condition is satisfied, i.e. every non-spacelike geodesic with tangent vector K^{a} contains a point at which $K_{[a}R_{b]cd[e}K_{f]}K^cK^d \neq 0$;

(4) there is an achronal set S such that $E^{-}(S)$ is compact.

This particular statement of the proposition derives from the two versions given in Hawking and Ellis 1973, pp. 266–267. A set such as S in (4) is called *past trapped*.

Lemma 2.9.5. Every LP spacetime possesses a past trapped set.

Proof. If (M, g_{ab}) is LP then there are points p and q such p $P(q)$, i.e. such that q $I^{-}(p)$ and every past endless causal curve through p intersects $J^{-}(q)$. Let λ be a past endless null geodesic curve with future endpoint p, and let $\mu=E^-(p)\cap\lambda$. Then $\mu\cap J^-(q)\ \ E^-(p)\cap J^-(q)\ \ \partial I^-(p)\cap J^-(q)\ \ \partial I^-(p)\cap I^-(p)=\ $ — which means that μ cannot be past endless, on pain of contradicting $p\ \ P(q)$. So μ has a past endpoint, s, say, as shown in figure 29. Now since μ is contained in the closed set λ, $s\ \ \lambda\ \ J^-(p)=E^-(p)\ \ I^-(p)$. But if $s\ \ I^-(p)$, then there would be points on μ arbitrarily close to s that were contained in $I^-(p)$ — which is impossible because μ is contained in $E^-(p)$. Hence $s\ \ E^-(p)$. Thus each generator of $E^-(p)$ is a null geodesic with future endpoint p and past endpoint in $E^-(p)$.

Now let $V\ \ T_p(M)$ be a future-pointing timelike vector, and define

$$Q(p)=\{U\ \ T_p(M)\mid g(U,U)=0, g(U,V)=1\}$$

$$Q^*(p)=\{V\ \ T_p(M)\mid V=cU, U\ \ Q(s), 0\leq c\leq f(U)\}$$

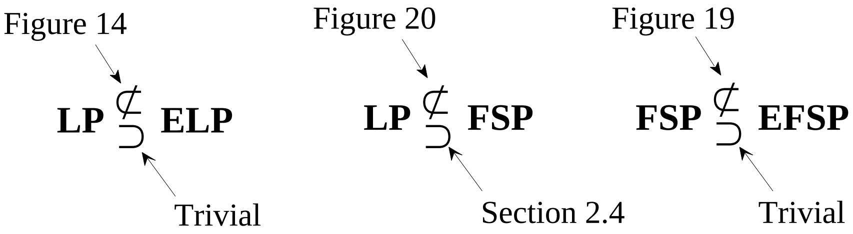
where $f(U)\ \ [0,\infty)$ is defined by $exp_p[f(U)U]$ is the endpoint of the null geodesic whose tangent vector is U. The aim is to show that $Q^*(p)$ is compact. (Where necessary think of the tangent spaces $T_p(M)$, $Q(p)$, and $Q^*(p)$ in terms of their corresponding Minkowski spacetimes.) Let $W_n=c_nU_n$ be a sequence of vectors in $Q^*(p)$, where U_n is in $Q(p)$. Since $Q(p)$ is compact, there is a limit point X of the U_ns in $Q(p)$. Choose a subsequence X'_n which converges to X and denote the corresponding subsequence of c_n by c'_n. Suppose that c'_n has no limit point in the interval $[0, f(X)]$. Then there exists a subsequence c''_n of c'_n with $c''_n\geq c$ for all n, where $c>f(X)$ is fixed and small enough so that $w=exp_p(cX)$ is defined. Since $w\ \ I^-(p)$, and the map exp_p is continuous, there is an integer n_0 such that $n\geq n_0$ implies $exp_p(c''_nX''_n)\ \ I^-(p)$. But this contradicts $c''_n\leq f(X''_n)$. So the sequence $c'_nX'_n$ has a limit point. In other words, every sequence in $Q^*(p)$ has a limit point in $Q^*(p)$, so $Q^*(p)$ is compact. Since the exponential map exp_p is continuous, the set $E^-(p)=exp_p\{Q^*(p)\}$ is compact. ♦

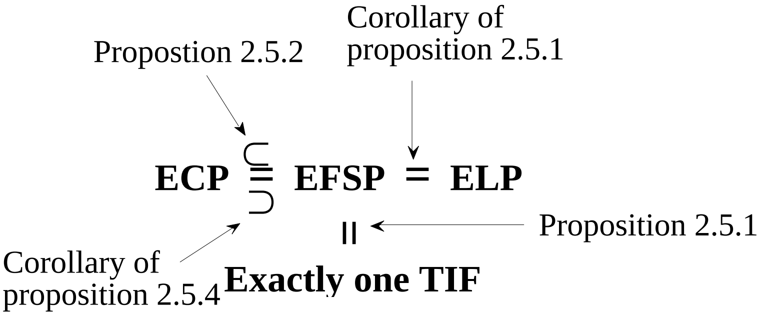
Lemma 2.9.5 applied to proposition 2.9.4 proves proposition 2.9.3. The obvious physical interpretation of this result is that prediction entails a singularity. However if one takes the view that closed timelike curves are not physically impossible, then the proposition could be interpreted as saying that prediction entails either a singularity or a closed timelike curve, or perhaps both. In any case, Corollaries 1 and 2 show that the anticipation region is free from such anomalies.

Proposition 2.9.3 obviously shows that ECP spacetimes (LP spacetimes) are normally singular. Although this reveals nothing about the nature of the expected singularity in ECP spacetimes, it is clear that it cannot be 'naked' because ECP spacetimes are necessarily globally hyperbolic (propositions 2.5.4 and 2.5.5), i.e. they satisfy strong cosmic censorship.

Summary and final remarks

Let LP denote ‘the class of LP spacetimes’, and let the classes of FSP, ELP, EFSP, ECP and ‘exactly one TIF’ spacetimes be denoted analogously. Further, let ‘ ’ denote ‘is a subclass of’, and ‘ ’ denote ‘is not a subclass of’. Then we have the following self-explanatory summary of our results.





In his 1977 paper, Geroch concludes that: ‘In physical terms, “predictions which can be verified directly arise only in a closed universe.” Why should this strange result follow from the basic principles of general relativity? Are there any other similar theorems about the domain of prediction?’ (p. 92). My response to these words is that prediction does not arise only in closed universe (section 2.3), but it does arise only in universes that possess a past trapped set (section 2.9).

Müller-Hoissen’s (1981) claim that FSP spacetimes are closed was also shown to be false. However, propositions 2.3.2 and 2.4.2 showed that Geroch’s claim and Müller-Hoissen’s claim will hold if an additional, arguably reasonable condition is imposed on spacetime.

Geroch also claimed that if a prediction could be made then there is always an achronal set from which the relevant data could be collected. Not true (section 2.3).

Section 2.5 revealed that the condition of being ELP is equivalent to the apparently much stronger condition of being ECP. This led to a dilemma that is worth repeating: *either the whole universe is predictable from any event or there exists at least one absolutely unpredictable event*. The equivalence of ECP and 'exactly one TIF' spacetimes highlighted the massive degree of constraint the ECP condition imposes. This was further emphasised by proposition 2.5.4, which showed that the past horismos of any point in a ECP spacetime is a compact Cauchy surface. As regards standard exact solutions of the Einstein equations: there seem to be none that admit even one predictable event. This makes the worry I describe in section 2.7, of being an inhabitant of an extendible ECP world, somewhat academic.

Hopefully light has been shed on the relationship between the concepts 'determinism' and 'prediction'. A first brush with these in the general relativity literature may leave a reader with the impression that a determined event is a predictable one and vice versa, so 'prediction' and 'determinism' are somehow equivalent. Sometimes this is due to a loose way of speaking, but sometimes there is a genuine belief that the two concepts are one. Karl Popper (1982) is perhaps the most famous advocate of this view. His belief that the world is unpredictable therefore leads him to claim that *ipso facto* the world is indeterministic. But in the manner of speaking adopted in this thesis and elsewhere (e.g. Earman 1986, chapter 2) I would say that the world might be unpredictable (because the information that determines the future can not be gathered up in time or at all) but still deterministic with respect to, say, a very early slice (the 'big bang') or the slice 'now'. The comparative richness of this account over Popper's illustrates how much better off we are for keeping the concept of determinism and predictability quite separate

In the same vein I note how misleading are the titles of two papers I referred to earlier, namely Müller-Hoissen's (1981) 'Determinism in Spacetime' and Budic and Sachs' (1976) 'Deterministic Spacetimes'. Both appear to be about relativistic *determinism* when in fact both are about relativistic *prediction*.

The conflation has also spawned a red-herring in the debate over the cosmic censorship hypothesis. One intuitive argument frequently advanced against naked singularities is that they upset determinism / predictability. So for example Penrose (1973), who advocates strong cosmic censorship (implies determinism), also talks about the '[t]he unpredictability entailed by naked singularities...' as being 'abhorrent' (p. 618). But we now see there are two distinct arguments here: (1) there is determinism, naked singularity upset determinism, *ergo* no naked singularities; (2) there is predictability, naked singularities destroy predictability, *ergo* no naked singularities. I have nothing new to say about (1), but (2) flops because the first premise is false — prediction is generally impossible.

Still on the relationship between singularities and prediction, it was shown in section 2.9 that in line with intuition prediction is possible only if the region between the prediction event and event predicted — the anticipation region — is non-singular. However, a less predictable result revealed that in any physically reasonable spacetime the region beyond the anticipation region must be singular. Prediction necessitates singularities. How curious.

3 Computability

3.1 Introduction

And the days are not long enough
And the nights are not long enough
And life slips by like a field mouse
Not shaking the grass.

EZRA POUND

The computer I am using to write these words performs a certain number of computational tasks or operations per second. I am not sure what the precise number is, but I know the Cray II performs many more operations in the same time and Babbage's difference engine far fewer. Yet despite these differences these machines share a fundamental property: the number of operations they perform in a finite span of time is itself finite. At first sight this observation might seem trivial because it is surely obvious that the smallest infinity — countable infinity, $\aleph_0$, — must stand beyond the reach of any computing device. (The days are not long enough.) If one is pressed to explain this further, then one might draw an analogy with special relativity by saying that computation rate can increase but never hit $\aleph_0$ in the same way that particle speeds can never reach the speed of light. But is this a good analogy? The fact that particles cannot break light speed is a *contingent* matter; the absence of a speed limit in Newtonian theory makes it clear that it could have been otherwise. So this raises a question: is the alleged ceiling on computation rate a contingent matter or is it true *a priori*? Putting the question another way: is it in some sense possible for a device to perform a *computational supertask*, i.e. an infinite number of computations in a finite span of time?

Virtually all mathematicians versed in current computability theory (i.e. Turing's theory) take it as axiomatic that the answer is no. A glance through the standard textbooks on the subject reveal that the possibility is not so much denied as not even considered. A finite time implies a finite number of computational tasks — period.

Yet the credibility of this axiomatic or *a priori* view is brought into question by the fact that computational supertasks do seem to be (at least) logically possible. One can imagine, for example, a mechanical device whose components accelerate so rapidly that it performs the first calculation in 1/2 second, the second in 1/4 a second, the third in a 1/8 second, and so on *ad infinitum*. A computational supertask, it seems, would be performed in one second flat. Then there is a mechanical device whose components shrink continuously to nothing in a finite time. If the velocity of the components remained constant, then a computational supertask again seems in the offing. The point of these strange scenarios is exactly that: they are strange but not obviously incoherent. They represent a way the world might have been but probably is not. In short they are, or at least seem to be, computer *possibilia*.

This simple observation may be seen as a hint that Turing's theory fails to cover the whole concept of computability. But actually much stronger evidence for a new concept is to be found from a quite different source, actually from a *Gedankenexperiment* in GR. Explaining this experiment and its revolutionary implications for computability is the object of this chapter.

I will proceed as follows. Building on a remarkable by idea Itimar Pitowsky, I give the conditions necessary for a spacetime to allow a supertask (section 3.2), which leads to the definition of a *Malament-Hogarth* (M-H) spacetime. Examples follow in section 3.3. Then in section 3.4 I create a worry that M-H spacetimes are in some sense logically flawed; but the worry is soothed. Section 3.5 provides some results concerning the global properties of M-H spacetimes, and also a result concerning the behaviour of light in such worlds. At this juncture attention is switched from supertasks to computability. The notion of a Turing machine is rehearsed, together with some results from Turing computability (section 3.7). A Turing machine operating in a M-H spacetime is shown to outperform an ordinary Turing machine by solving, among other things, the halting problem (section 3.8). A series of more powerful machines can decide arithmetic (section 3.9). All these machines are axiomatised in the manner of Turing's hypothetical machine, which allows us to assess their power (section 3.12). Finally, in section 3.13, I advance a new theory of computability based on these findings.

3.2 Pitowsky's idea

To see a World in a Grain of Sand
And a Heaven in a Wild Flower,
Hold Infinity in the palm of your hand
And Eternity in an hour.

WILLIAM BLAKE, *Auguries of Innocence*

For an effective partnership, a computer user and her computer should remain quite close to each other — in spatio-temporal terms, that is. But Pitowsky (1990) pointed out that if this relationship is broken then a tantalising prospect arises. For if the computer follows a different worldline, its clock will tick at a different rate to that of the computer user's clock; and so perhaps, reasoned Pitowsky, an extreme case could be arranged in which the rates are such that the finite proper time as measured by the computer user 'corresponds' to an infinite proper time as measured by the computer. In this case, and granting also that the computer can always signal to the computer user, the computer user will take only a finite time to view the eternity of the computer's life and the fruits of its more than Herculean labours.

Is this possible in a relativistic spacetime? Pitowsky answered positively and claimed, moreover, that the scenario could even be enacted in Minkowski spacetime. It is enlightening to see why this is not so. In Hogarth (1992) I illustrate the point with a story, set in Minkowski spacetime, about an immortal computer HAL and mortal computer user Dave. Dave is a passionate number theorist whose desire to know the truth of the Goldbach conjecture borders on the obsessional. Desperate for an answer, he takes his problem to the greatest computer on earth, HAL, who is happy to sacrifice her eternal life to this great question. Dave suggests to HAL that she should begin by first checking if 2 is the sum of two primes, then 4, then 6, and so on *ad infinitum*; and if at any stage she finds a counter-example in this way then she must report it to Dave immediately. Bidding each other adieu, Dave and HAL then pursue different worldlines though spacetime, as depicted in figure 30.

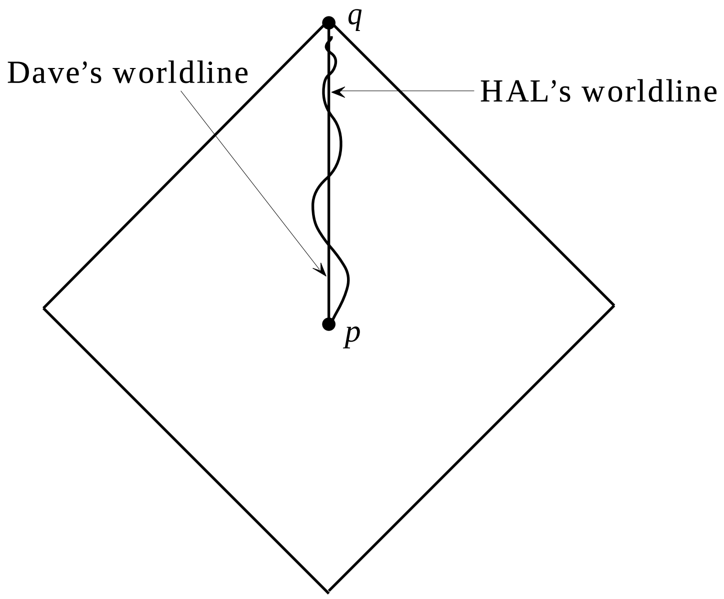


Figure 30. Conformal diagram of Minkowski spacetime. HAL follows as inertial worldline, while Dave oscillates to-and-fro about HAL's worldline on a worldline that approaches a null zig-zag line exponentially quickly and is thus of finite length.

The key question is: will Dave come to know the truth value of the Goldbach conjecture? There are two cases to consider. First suppose that the Goldbach conjecture is false. Then HAL will eventually discover a counter-example and Dave will be informed. So far so good. On the other hand, if the Goldbach conjecture is true then Dave will never know that fact *within* spacetime for the point *q* in figure 30 is a point at infinity, not a normal spacetime point.

This uncovers a fundamental flaw in Pitowsky's conception of what is required of spacetime to enable a computational supertask. The difficulty is that at any event Dave can access only a finite part of HAL's life. It is true that this provides Dave with a quantitative advantage over checking for counter-examples without HAL, but in terms of guaranteeing sure knowledge about the truth / falsity of conjectures like the Goldbach conjecture, the advantage is nil. The task is impressive but not super. [17]

What then is required of spacetime for the implementation of a genuine computational supertask? Figure 30 provides the clue. Instead of the point *q* being at infinity, it should be a normal spacetime point. In that way Dave would reach a real event from which the whole of HAL's eternal life would be visible. I will follow

---

[17] This is also true of David Deutch's so-called quantum computer: it is merely the functional equivalent of a very fast conventional computer. See Penrose 1994, p. 356.

Earman and Norton (1993) by calling a spacetime with this property *Malament-Hogarth* (M-H).[18]

Tentative definition. A spacetime (M, g_{ab}) is *Malament-Hogarth* just when there is a future endless timelike curve λ with past endpoint, and a future endless timelike curve μ with the same past endpoint as λ and future endpoint q; such that

$$(1) \int_{\lambda} ds^2 = \infty, \quad (2) \int_{\mu} ds^2 \text{ is finite}, \qquad (3)\, \lambda \quad J^{-}(q).$$

The curves λ and μ represent, respectively, the worldlines of HAL and Dave; q is an event in Dave's life. That μ and λ share the same past endpoint reflects the fact that HAL and Dave are initially together. Conditions (1) and (2) reflect the longevity of the two observers: HAL lives forever, while Dave lives for only a finite time. Condition (3) ensures that HAL can always signal to Dave and, moreover, that every such signal can be sent to Dave at q.

In fact the definition of a M-H spacetime admits a simpler formulation.

Definition 3.2.1. A spacetime (M, g_{ab}) is *Malament-Hogarth* just when there is a future endless timelike curve $\lambda \quad M$ with past endpoint and a point $q \quad M$ such that

$$\text{(i)} \int_{\lambda} ds^2 = \infty, \quad \text{(ii)}\, \lambda \quad J^{-}(q).$$

(Hereafter, the symbols 'q' and 'λ' are assumed to have the properties they have in definition 3.2.1. Following Earman and Norton (1993) I will sometimes refer to q as the *M-H event*. I will also speak of a 'λ-curve'.)

This follows because each spacetime that satisfies the conditions of definition 3.2.1 automatically admits a curve with the properties of μ in the Tentative definition, i.e., a timelike curve of finite length which has the same past endpoint, p say, as λ, and which passes through q. To see this, notice that since λ is timelike (ii) implies that p can be joined to q by a timelike curve, ν, say. But any timelike curve between two events must have finite length (*proof*: since the curve is a compact set, it can be

---

[18] In Hogarth 1992 I called this class of spacetimes *Pitowsky*. But Earman and Norton use this name for the class of spacetimes Pitowsky thought (mistakenly) would be suitable for the implementation of a computational supertask. David Malament and I arrived at the Malament-Hogarth definition independently.

covered by a finite number of convex normal neighbourhoods such that it has length less that some fixed ε in each such neighbourhood), so ν is the required curve.

Definition 3.2.1 states that a spacetime is M-H if it admits a point to the future of a curve with past endpoint and infinite length.

3.3 Examples of Malament-Hogarth spacetimes

A particularly simple and causally well-behaved M-H spacetime is given by Earman and Norton (1993), and is the depicted in figure 31. To create the example, start with Minkowski spacetime $(\mathbb{R}^4, \eta_{ab})$ and choose a scalar field Ω on M such that $\Omega=1$ outside a compact set $C \subset M$ and Ω tends rapidly to infinity as a point $r \in C$ is approached. The spacetime $(\mathbb{R}^4 - r, \Omega^2\eta_{ab})$ is manifestly M-H.

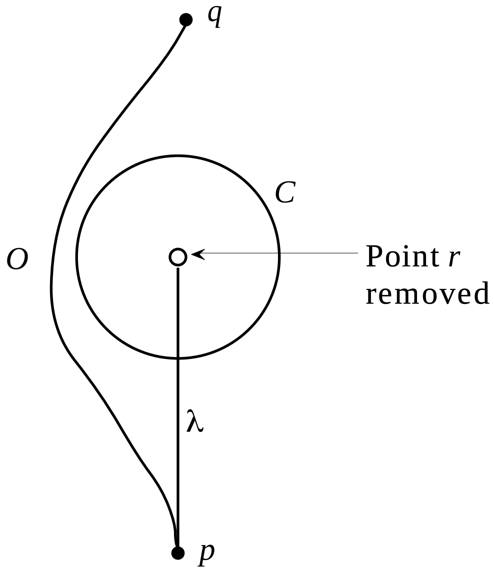


Figure 31. A toy Malament-Hogarth spacetime.

This provides a concrete arena in which to perform the advertised supertask. From a point $p \in \lambda$ launch a computer (e.g. HAL) along λ that is primed first to check if 2 is the sum of two primes, then likewise to check 4, then 6, and so on, *ad infinitum*. The computer is also primed to signal to q if and only if it finds a counter-example to the conjecture, and then to halt operations. Since an observer O (e.g. Dave) can travel from p to q in a finite span of proper time, he can discover the truth of the conjecture before the day's over. (The days are long enough!) Fermat's last theorem is cracked in a similar fashion.

(This spacetime has some intriguing properties that are worth mentioning. First, although the region inside C *appears* quite small, it is in fact as large as the complement of C. This counter-intuitive aspect is made even more stark by observing that while the surface area of C is finite the volume it 'contains' is infinite. In a sense then C and its contents represent an extreme version of Doctor Who's *Tardis*. Recall that this curiosity has the exterior of a telephone box but an interior which is larger than a telephone box interior. This means that objects larger than a telephone box can be accommodated within the Tardis, but they cannot be brought whole through the door: they must first be dismantled and then reassembled on the inside. This is also true in essence of the world in figure 31. Objects whose spatial dimensions exceed C's spatial dimensions cannot pass whole into C. So the boundary of C acts as a spacetime filter, letting through only stuff of less than a critical radius; anything else either avoids C completely or is partly in C and partly out. Again, a process of dismantling and reassembling is required.)

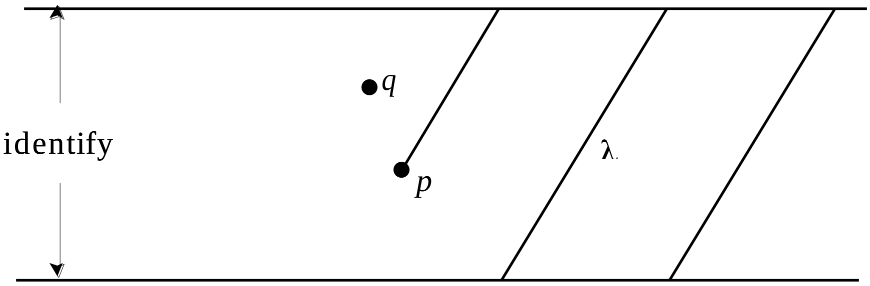


Figure 32. A causally vicious Malament-Hogarth spacetime

Another toy example is afforded by the cylindrical spacetime formed by indentifying the edges of a temporally finite strip of 2-dimensional Minkowski spacetime. It is clear that the M-H property is satisfied, but unlike the example above there are closed timelike curves, actually through every point. Whether this property alone makes this spacetime beyond the pale is something I touched on in the previous chapter. Here I simply note that those in the 'physical' school may find the example significant and those in the 'unphysical' school can look forward to the subsequent examples which are all causally well-behaved. In any case, nothing with regard to the kind of computability I have in mind will turn on causality issues, so hereafter I assume that spacetime is causally well-behaved in the sense of being strongly causal.

Anti-de Sitter spacetime is M-H (Hogarth 1992).[19] This spacetime, which is stably causal and *a fortiori* strongly causal, can be covered (Hawking and Ellis 1973, p. 131) by a single ‘spherical polar’ coordinate system (t,r,θ,ϕ), in which case the line element takes the form

$$ds^2 = cosh^2 r\ dt^2 - dr^2 - sinh^2 r\ (d\theta^2 + sin^2\theta\ d\phi^2).$$

Restricting attention to the plane $\theta=\phi=0$, gives a line element

$$ds^2 = cosh^2 r\ dt^2 - dr^2.$$

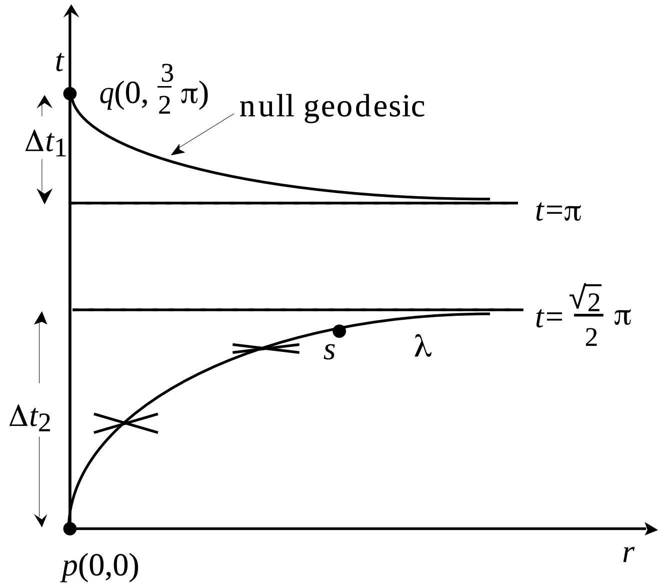


Figure 33. Anti-de Sitter spacetime is Malament-Hogarth.

In keeping with the notation of definition 3.2.1, let p and q be the points (0, 0), (0, $\frac{3}{2}\pi$), respectively, and let λ be the future endless timelike curve with past endpoint p whose tangent vector satisfies

$$\frac{dt}{dr} = \frac{2^{1/2}}{cosh\ r}\ .$$

The aim is to show that p, q, and λ satisfy the conditions (i) and (ii) of definition 3.2.1. To satisfy (i) it must be shown that the length of λ , $L(\lambda)$, is infinite. We have

---

[19] David Malament suggested this example to me.

$$L(\lambda) = \int_0^\infty \frac{ds}{dr}\,dr = \int_0^\infty (\cosh^2 r\,(\frac{dt}{dr})^2 - 1)^{1/2}\,dr = \int_0^\infty dr = \infty.$$

Next, to satisfy (ii), we need to show that $\lambda \;\; J^-(q)$. Setting ds=0, shows that the past null geodesic through q is given by $dt/dr = 1/\cosh r$ (see figure 33). The total decrease in t along this geodesic, Δt_1, is

$$\Delta t_1 = \int_0^\infty \frac{dt}{dr}\,dr = \int_0^\infty \frac{1}{\cosh r}\,dr = [2\,\arctan(e^r)]_0^\infty = \frac{\pi}{2}$$

Thus every null geodesic with future endpoint q is contained in the set $\{(t,r) \mid \pi < t \le \frac{3}{2}\pi,\ 0 \le r < \infty\}$. For any point $s=s(t,r)$ with $t<\pi$, the timelike curve through s given by $r=constant$ intersects the null geodesic which passes through q. Hence $s \;\; J^-(q)$. But this applies to every point $s \;\; \lambda$ since the t coordinate along λ never exceeds

$$\Delta t_2 = \int_0^\infty \frac{dt}{dr}\,dr = \int_0^\infty \frac{\sqrt{2}}{\cosh r}\,dr = \frac{\sqrt{2}}{2}\,\pi \;\; (<\pi).$$

(ii) is therefore satified and so anti-de Sitter spacetime is M-H.

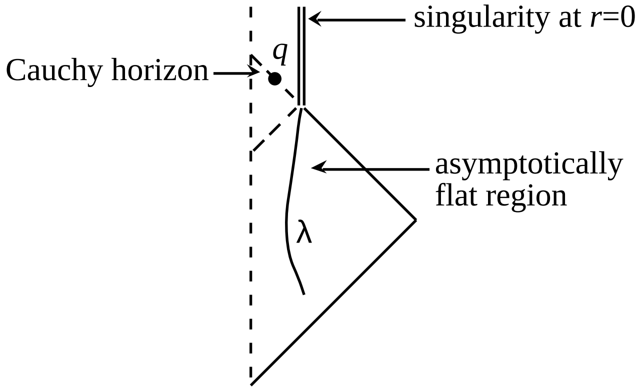


Figure 34. The Reissner-Nordström solution is Malament-Hogarth.

The maximally extended Reissner-Nordström solution is also M-H. This can be seen from the spacetime's Carter-Penrose diagram depicted in figure 34. The future endless timelike curve λ is contained in the asymptotically flat region and is of infinite length. (Of course the fact that λ is infinitely long cannot be inferred from the Carter-Penrose diagram; rather it follows because the region containing λ is asymptotically Minkowski spacetime.) The spacetime is M-H because λ lies to the past of *q*, a point on the Cauchy horizon.

3.4 Do Malament-Hogarth spacetimes yield paradox?

Supertasks and paradoxes are known associates. But are the paradoxes real or do they dissolve under scrutiny? Consider the case of Zeno. He thought he had found a proof that continuous motion was impossible because for a body to move from point *A* to point *B* required the body to traverse an infinite number of distinct spatial intervals between those two points. The orthodox reply, however, is that this does not represent a genuine supertask because the runner does not perform an infinite number of distinct 'acts'. And even if he is made to do so, say by pausing between each interval and planting a flag, then there is no paradox in the sense of an outright contradiction. The situation is simply extraordinary, not incoherent (See Grünbaum 1968).

Thompson's Lamp

More relevant to our discussion than Zeno's paradox is Thompson's lamp (Thompson (1954-55)). Imagine a lamp that is switched on at t=0, then is switched off at t=1/2s, on at t=1/4s, and so on. At t=1, is the lamp on or off? Thompson argued that it could not be off because to be so would entail it being off throughout (t',1], for some t' — but that is impossible because in *any* open interval in (0,1), there is a time when the lamp in on. A similar argument proves that the lamp cannot be on. Hence the declaration of paradox.

Now the worry for us derives from the fact that the Thompson lamp scenario maps neatly into any M-H world. Have the lamp travel on the λ-curve, switching itself alternately off and on at each unit interval of λ's proper time. As Dave travels towards the M-H event, he can witness each such switching. The question is: at the M-H event, does Dave perceive the lamp to be on or off? Paradox looms again. Have we unearthed a conceptual flaw in M-H worlds?

No. For the Thompson Lamp paradox in its classical and M-H form represents, like Zeno's paradox, only a strange situation, not a contradictory one. The classical version was first resolved by Benacerraf (1962). He pointed out that since the time t=1 is strictly outside the time interval [0,1) in which the lamp's state is supposed to change, the lamp's state prior to t=1 does not logically force *any* particular state *at* t=1; the state could be on or off or exploding or whatever. (Grünbaum (1968) was thus able to construct two lamp scenarios, both conforming to Thompson specifications, but in one case the lamp was on at t=1, while in the other it was off at t=1.) This is the crux of the matter.

Not everyone, however, is convinced by the combined arguments of Benacerraf and Grünbaum. Ray (1991, chapter 1), for example, maintains that paradox still threatens. It is instructive to see how his argument fails (this is touched on by Clifton and Hogarth (1993)).

He considers two identical Thompson lamps that are switched on and off in parallel. From this he reasons that we are ‘left with the unsatisfactory conclusion that two [lamps] always in step during an infinite sequence of tasks may be out of step immediately after the sequence has “ended”’. And why unsatisfactory? ‘...we need offer no reason other than considerations of symmetry between the two lamps to justify the belief that the lamps will be in the same state...’. Upshot: ‘If we were to provide even a sketch explanation for the broken symmetry, we would be supplying some kind of link between the lamps in operation during the infinite task and the lamps at t=1’.

But Ray is not taking Benacerraf seriously. If we persist in demanding the lamps be in the same state at t=1 — be it on, off, or whatever — then our belief that these states be identical ‘by symmetry’ must appeal to (some form of) determinism of later states by earlier ones. That supplies the missing link. The point is that sheer conceptual analysis about what the lamps do during [0,1) cannot decide what their states will be at t=1 independent of further specification of how they operate beyond [0,1). And if by asking what their states will be at t=1, we are implicitly demanding to know the final state they were put into just before that moment, then we shouldn’t be asking — there is no such state.

The Thompson Lamp paradox is therefore really only a harmless puzzle. But what about its M-H counterpart? A harmless puzzle also, for exactly the same argument shows that the state of the M-H event is not logically determined by the lamp’s history. By way of illustrating this point, consider the scenario depicted in figure 35. At any event prior to the M-H point Dave can see the lamp via the past null geodesic through that event. It is clear that the order of emission and reception match. His record therefore shows that the lamp is on, then off, then on, and so on. But — and this is the point — at the M-H event there is no such null geodesic from the lamp to fix what he sees (if such a signal was sent by HAL at an event r HAL could not communicate with Dave beyond r without destroying the emission/reception matching property inherent in the set-up).

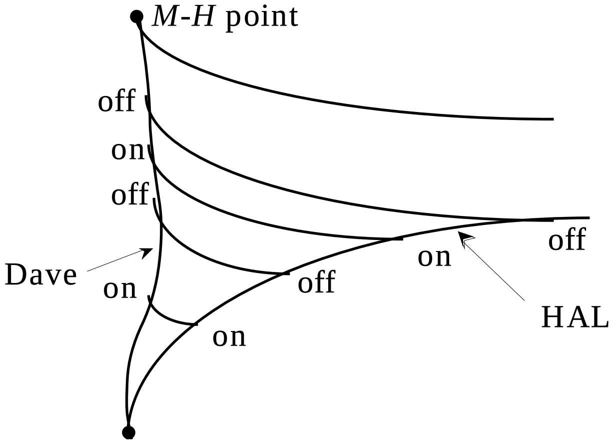


Figure 35. Dave looks on while HAL switches the lamp alternately on and off.

So much for Thompson's Lamp. Now I want to discuss another strange scenario.

Littlewood's balls

'An infinity paradox. Balls numbered 1,2,... (or for the mathematician the numbers themselves) are put into a box as follows. At 1 minute to noon the numbers 1 to 10 are put in, and the number 1 is taken out. At 1/2 minute to noon numbers 11 to 20 are put in and number 2 is taken out. At 1/3 minute to noon 21 to 30 in and 3 out; and so on. How many are in the box at noon? The answer is none: any selected number, e.g. 106, is absent, having been taken out at the 106th operation.' (Littlewood 1986, p. 26.)

The way this can be mapped into an M-H arena is clear: HAL has the box and the balls, while Dave receives signals from HAL about how many balls are in the box at that moment. The question is: at the M-H event, how many balls does Dave perceive to be in the box? But before addressing this question, let us return to Littlewood.

The first thing to note is that *if* Littlewood is right and the box is empty at noon, then the situation is not paradoxical but merely counter-intuitive. There are not two conflicting arguments in play, as there were with Thompson's Lamp. There is one argument and that asserts the box is empty at noon.

Littlewood, however, is not right. In the light of the remarks above it is clear that the box is not necessarily empty at noon, for *that* event is simply not determined by the previous operations. The question Littlewood poses is simply underdescribed.

And with regard to the corresponding M-H scenario, as depicted in figure 36, it is also clear that the situation is exactly analogous to the Thompson lamp scenario. In the lead up to the M-H event, Dave infers how many balls there are in the box from what he sees (9, 18, 27,...), but whatever he sees at the M-H event is not determined by these events.

Of course it is true that the M-H event could be determined by the balls if there was an appropriate physical constraint in place. For example if it was deemed that the number of balls visible at the M-H event must correspond to the 'limit' of the ball-counts in the lead up to that event, then Dave would perceive an infinite number of balls. On the other hand if it is deemed that at the M-H event no ball that has previously exited the box can be seen then the box would appear empty. Which condition is the more realistic is an interesting question but a question of *physics*: and so the outcome of that investigation will not alter the fact that there is no genuine *logical* paradox in the offing. For that reason I will not pursue the matter further (Earman and Norton 1994 argue for the empty case; see also Ross 1988).

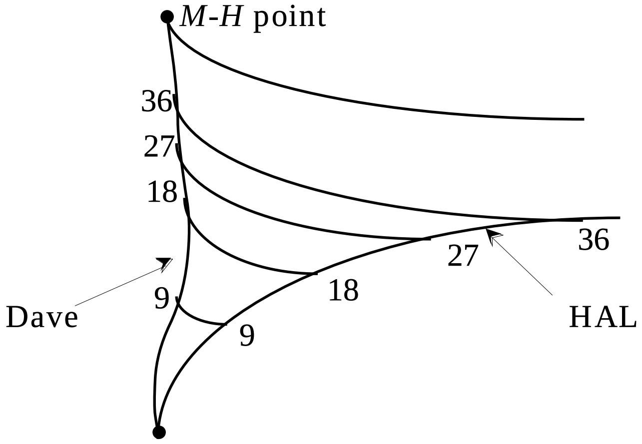


Figure 36.

The conclusion is that the box and ball problem also fails to reveal any chinks in the logic of M-H spacetimes.

To recap this section, I have shown that at least one alleged paradox, the Thompson Lamp, that threatened to undermine M-H spacetimes is actually harmless; and that the Littlewood box problem could be similarly defused. Indeed, these two puzzles in their original form are resolved (or dissolved) more easily in the light of their M-H counterparts. M-H spacetimes have thus shown themselves to be sharp tools for dissecting some classical puzzles. Of course, all these findings do not constitute a proof that M-H spacetimes are consistent, which was the initial worry. But pending evidence to the contrary we should feel justified in proceeding with equanimity.

3.5 The structure and physicalness of Malament-Hogarth spacetimes

That the unusual M-H property forces some particular global constraints on spacetime is to be expected. I will now give two results along these lines.

Proposition 3.5.1. Every Malament-Hogarth spacetime is not globally hyperbolic.

Proof: Suppose (M, g_{ab}) is a globally hyperbolic spacetime, and suppose further that there is a timelike curve λ and a point q such that $\lambda \quad J^-(q)$. (It will be shown that the length of λ must be finite — thus preventing (M, g_{ab}) being M-H.) Since strong causality holds on (M, g_{ab}) , the set $J^-(q) \cap J^+(p)$ has an open cover $\{U_\alpha\}$, where each U_α is a convex normal neighbourhood with compact closure such that λ enters each U_α at most once. From this cover construct another cover $\{V_\beta\}$ of $J^-(q) \cap J^+(p)$, where each $V_\beta \quad U_\alpha$ for some α and such that, for each V_β with $V_\beta \cap \lambda \neq \quad$, the length of λ in V_β does not exceed some fixed $\varepsilon > 0$. Now, since (M, g_{ab}) is globally hyperbolic, $J^-(q) \cap J+(p)$ is compact, which implies that $\{V_\beta\}$ has a finite subcover. This cover contains λ since $\lambda \quad J^-(q) \cap J^+(p)$. Therefore λ can be covered with a finite number of the V_βs, each of which contains a segment of λ of finite length. Thus λ has finite length and (M, g_{ab}) is therefore not M-H.♦

The first statement of this proposition was given by Hogarth (1992), actually with the proof above. Earman and Norton (1993) give two alternative proofs, one of which they attribute to David Malament. Both these proofs and the one above use the fact that λ is infinitely long; but in fact a quick proof can be given that uses only the fact that λ is future endless. This depends upon following:

Lemma 3.5.2 (Hawking and Ellis 1974, proposition 6.4.7). If the strong causality condition holds on a compact set S, there can be no future-endless non-spacelike curve totally future imprisoned in S.

(A curve μ is *totally imprisoned* in a compact set R if μ enters and remains in R.)

Alternative proof of proposition 3.5.1. Suppose (M, g_{ab}) is globally hyperbolic. Then for any two given points p and q, $J^-(q) \cap J^+(p)$ is compact, so by lemma 3.5.2 this set cannot contain a future endless curve. Thus (M, g_{ab}) is not M-H.

Corollary. If for a spacetime (M, g_{ab}) $N \neq \quad$ (i.e. there is a naked singularity), then (M, g_{ab}) is not globally hyperbolic (i.e. strong cosmic censorship is violated.)

Proposition 3.5.1 shows that every M-H spacetime must violate strong cosmic censorship (as defined in chapter 1). It is also clear that every M-H spacetime must possess a naked singularity (i.e. N≠) because by definition λ lies in q's past. This pair of results may give the impression that M-H spacetime are singular in some fairly strong sense. But notice that not every M-H spacetime is singular even in the sense of being geodesically incomplete: think of anti-de Sitter spacetime. Thus although M-H spacetimes do necessarily violate strong cosmic censorship and possess a form of naked singularity, some M-H spacetimes are singular only in a very weak sense.

An immediate corollary of proposition 3.5.1 is that Minkowski spacetime, the standard Robertson-Walker models, Taub space, the Einstein static universe, the Schwarzschild black hole and Schwarzschild white hole are not M-H because each of these spacetime is globally hyperbolic. This puts paid to an idea suggested to me a few times, that a supertask could be performed by having an observer fall into a black hole.

Another corollary (Earman and Norton 1993, lemma 3) is this: if (M, g_{ab}) is M-H and S M is any achronal set such that λ   D(S), then the M-H point q must lie outside $int\{(D(S)\}$. For otherwise $int\{D(S)\}$ would be a M-H spacetime, which is impossible because $int\{D(S)\}$ is necessarily globally hyperbolic (Hawking and Ellis 1973, proposition 6.6.3). Physically speaking this result means that any achronal set that determines λ cannot also determine q. Earman and Norton infer from this that Dave (to use my character) therefore has a serious problem in knowing whether the signal received at the M-H event q originated from HAL or from beyond the Cauchy horizon. No achronal set can possibly hold the answer.

I agree this is a problem. But it is wrong to think, as Earman and Norton seem to, that the problem of confirming the source of a computer's signal is something special to M-H computation. The problem is endemic; one might even say it is an *intrinsic feature* of computation. To see why, suppose that Dave is using HAL as a conventional computer to perform some calculation. If all goes well then Dave will eventually receive a signal from HAL. But to know for sure that this signal came from HAL would presumably require (1) that the event at which the signal is received is determined by a past state of HAL; (2) that Dave can access that past state; (3) that Dave can compute the evolution of that past state. (These conditions are redolent of the conditions required for prediction given in the previous chapter.) This is a very tall order, not least because (3) would involve Dave calculating *en passant* the very calculation he set HAL to perform!

What this means is that if computers are to aid rather than encumber, then we must not attempt to confirm the origin of the signals we receive from them. Seen this way, Earman and Norton's result is interesting but not particularly significant.

The following simple result (which has not appeared in the literature) shows that in a precise sense every M-H spacetime cannot be spatially closed throughout its history.

Proposition 3.5.2. Every Malament-Hogarth spacetime admits a non-compact slice.

Proof. Let p, q, and λ be as usual. If $\partial I^{-}(\lambda)=clos\{I^{-}(\lambda)\}-I^{-}(\lambda)\}$ were empty then $I^{-}(\lambda)$ must be open and closed, which implies that either $I^{-}(\lambda)=$ or $I^{-}(\lambda)=M$. But the first statement cannot hold because λ $I^{-}(\lambda)$, and the second statement cannot hold because by hypothesis λ $I^{-}(p)$ implies $I^{-}(\lambda)$ $I^{-}(p)$ implies p $I^{-}(\lambda)$, where the last implication follows because stable causality implies that there are no closed timelike curves. $\partial I^{-}(\lambda)$ is therefore non-empty and so by proposition 6.3.1 in Hawking and Ellis 1973, it is a slice generated by null geodesics that either have no future endpoint or intersect λ at a future endpoint. But the latter is impossible because λ future endless and timelike implies λ $I^{-}(\lambda)$ implies $\lambda\cap\partial I^{-}(\lambda)=$. Thus each generator of $\partial I^{-}(\lambda)$ therefore has no future endpoint. Let μ be a maximally extended generator of $\partial I^{-}(\lambda)$. Notice that since μ is closed, any sequence of points in μ that converges must converge to a point in μ. Now let x_1, x_2, x_3,... be a sequence of points in μ such that for any given point x μ, there is an n such that for every $m>n$ x_m lies to the future of x (this sequence is possible because μ is future endless). By construction this sequence clearly admits no convergent subsequence in μ, so by the preceding remark it has no limit in μ $\partial I^{-}(\lambda)$. Thus $\partial I^{-}(\lambda)$ is non–compact.♦

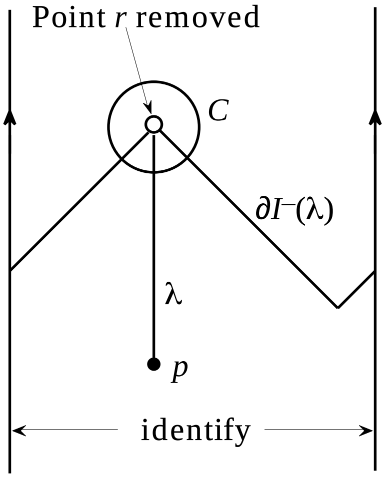


Figure 37. An ‘almost closed’ Malament-Hogarth spacetime. The spacetime is formed by first taking a spatially finite segment of two dimensional Minkowski spacetime and identifying the edges, removing a point r, and finally introducing an infinite conformal scaling on the region C à la figure 31. The slice $\partial I^{-}(\lambda)$ is non-compact.

Corollary. Every nakedly singular (strongly causal) spacetime admits a non-compact slice.

Proof. The proof of proposition 3.5.2 applies directly.♦

Proposition 3.5.2 provides another demonstration — the other being proposition 3.5.1 — that Taub spacetime and the Robertson-Walker $k=1$ are not M-H, since both these spacetimes possess only compact slices (actually S^3).

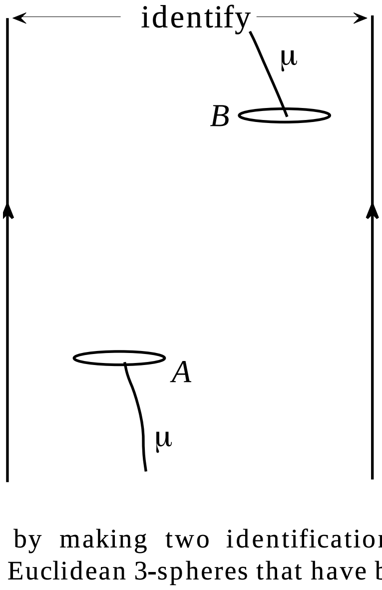


Figure 38. A spacetime formed by making two identifications in 4-dimensional Minkowski spacetime. A and B represent two Euclidean 3-spheres that have been removed, and the upper edge of the spacetime where it used to meet A has been identified with the lower part of the spacetime where it used to meet B. Observer μ is shown passing through the resultant wormhole. By proposition 3.5.2 the spacetime cannot be M-H.

One might envisage a way of converting an always spatially closed universe into an M-H spacetime by adding a ‘wormhole’ (as described in Morris and Thorne 1988), the idea being that these structures might provide a cosmic shortcut of the kind Dave needs to reach an M-H event. But proposition 3.5.2 seems to put paid to this plan because wormholes generally connect spatial sections rather than ‘open’ them out. The resulting spacetime would therefore also be always spatially closed and consequently not M-H. See Figure 38.

3.6 Light in Malament-Hogarth Spacetimes

Light, light,
the visible reminder of invisible light.

T. S. ELIOT, *The Rock, Chorus IV*

Propositions 3.5.1 and 3.5.3 provide information about the geometrical structure of M-H spacetimes, but now we switch attention to the behaviour of physical light in the universes underpinned by M-H spacetimes. This issue is pertinent because the success of the computational supertask relies upon HAL communicating information to Dave — and the canonical vehicle of communication in relativity is light.

Consider again the Reissner-Nordström solution, as depicted in figure 34. This is M-H, as we saw, and we know that the M-H point *q* must lie on or to the future of the Cauchy horizon, in accordance with a corollary of proposition 3.5.1. Now it can be shown by an explicit (and highly non-trivial) calculation that a photon that tries to approach this horizon becomes more and more blue shifted and thus more and more energetic (see Wald 1984, chapter 12). At the horizon itself, the energy is unbounded and so one cannot expect that the advertised geometry will be the actual one. In short, the horizon is unstable. (Recall the evidence for the CCH.) This observation would be incidental to our discussion were it not for the fact that this behaviour seems to be quite widespread among M-H spacetimes. There is actually an intuitive reason why this is so. During HAL’s infinite life, her light source will vibrate an infinite number of times. Dave will agree that an infinite number of vibrations have occurred, but must acknowledge that this happened within a finite span of his proper time. This it would seem can only happen if HAL’s source is seen by Dave to have an unbounded frequency.

But, as Earman and Norton (1993) point out, this argument is problematic because it depends upon the frequency of HAL’s source as inferred by Dave, when in fact what matters is the frequency of the photon that reaches Dave. Earman and Norton compute this frequency and use the result to prove the following (in my notation):

Lemma 3.6.1 (Earman and Norton 1993). Let (M, g_{ab}) be a Malament-Hogarth spacetime containing a timelike curve λ_H and another curve λ_D from point p to point q such that

$$\int_{\lambda_H} ds = \infty, \quad \int_{\lambda_D} ds < \infty, \text{ and } \lambda_H \quad I^-(q).$$

Suppose that the family of null geodesics from λ_H to λ_D forms a two-dimensional integral submanifold in which the emission from λ_H matches the order of reception at λ_D. If the photon frequency ω_H as measured by the sender λ_H is constant, then the time-integrated photon frequency $\int^{q_*} \omega_D \, d\tau$ as measured by the receiver λ_D diverges as q_* approaches q.

The proof is given in *ibid.* and is omitted here.

Photons of equal frequency are emitted by HAL at equal time intervals, and the time integral of these frequencies as measured by Dave becomes unbounded as Dave approaches q. The general message then is that M-H spacetimes entail infinite blue shifts. But a qualification is in order regarding the premise that the order of emission and reception must match. Typically this premise does seems to be met, but it is not met by, for example, the rolled-up Minkowski spacetime example depicted in figure 32 (because all the signals get jumbled up). And even in causally well-behaved M-H spacetimes there is no clear reason why it *must* hold.

On the other hand if the premises of lemma 3.6.1 do hold then the situation is even more dangerous than it might first appear. For even if HAL doesn't deliberately send a signal, the natural vibrations of her source will provide the radiation necessary for a blue-shift disaster. This of course bodes ill for the physicalness of M-H spacetimes, but it is far too soon to conclude from this meagre analysis that M-H spacetimes are beyond the physical pale.

(Idealised M-H scenarios which conform to some of our physical laws might be considered by switching the lights off. That is to say, one might consider M-H universes sans radiation. After all it would seem that HAL and Dave could in some sense operate and communicate *as purely mechanical devices* (in a kind of billiard ball model). The point of the exercise, of course, is that these worlds do not possess the seeds of blue-shift destruction.)

3.7 Orthodox Computability

‘Moreover it is absolutely impossible that anybody who understands the question and knows Turing’s definition should decide for a different concept.’

HAO WANG

Experiment escorts us last —
His pungent company
Will not allow an Axiom
An Opportunity

EMILY DICKINSON

I now turn from spacetime theory to discuss the theory of computability. I will also speak about the necessarily vaguer *concept of computability* (written *Computability*), which is just the notion the theory is trying to capture. What is this theory? And, more fundamentally, what is it concerned with? In abstract terms it is concerned, or seems to be concerned, with the ‘ideal computing device’ (or ‘effective computing device’ or ‘computing agent’ or just ‘computer’), what it is exactly, and the extent of its powers. The theory began to take shape in the early 1930s, especially with Gödel’s 1931 paper on primitive recursion. Then in 1936 four seminal papers were published — by Church, Kleene, Turing, and Post — that each advanced a precise definition of computability. At first, all these proposals appeared to be quite different from each other, but a detailed analysis soon showed them to be equivalent. This confluence of ideas is often taken as strong evidence that the proposed definition is indeed the right one.

In my analysis and subsequent criticism of this view, I will take one of these theories, Turing’s, as representative of the others. Computability for our purposes can therefore be described as the theory of Turing’s hypothetical computing device, what it can and cannot compute. The ultimate aim of this chapter is to show that this particular theory (and *ipso facto* those of Church, Kleene and Post) does not adequately capture Computability. It will emerge that Turing’s theory is not so much wrong as limited, for it describes only one aspect of Computability (just as Euclidean geometry describes only one aspect of the concept of Geometry).

Turing's idea

Turing sought to characterise the most general kind of computer. The construction he arrived at consists of two parts. One part is a device which operates in an entirely mechanistic way, like a clockwork watch. The other part is an infinitely long paper tape which is marked off along its length into cells. At any one time the device is scanning one cell. Each and every operation of the device takes place in (a fixed) unit interval time. The device is capable of four kinds of operation:

(1) it can write a '1' on the cell it is examining.

(2) it can erase the cell it is examining and thus make it blank.

(3) it can move to the next cell on the right.

(4) it can move to the next cell on the left.

The device can take on one of a *finite* number of states, denoted q_1, q_2, etc. The states will in general change as the device operates. Let 1, B, R and L denote the basic operations (1), (2), (3), and (4). Then the operation of the device can be given as a finite set of quadruples which consist (in order): (i) an internal state, (ii) a possible condition of a tape-cell, (iii) an operation, (iv) an internal state. A quaduple of the form i, ii, iii, iv expresses that given (i) and (ii), the device should then perform (iii) and enter the state in (iv).

The device together with its set of quadruples is known as a Turing machine (TM).[20] Given a particular set of quadruples (*instruction set*) and an initial state, the TM will carry out the instructions in a set way, which will either end in a finite number of steps or will never stop. So for example suppose the tape is a finite string of 1s, and the instruction set $\{q_1,q_2\}$ possessed by the TM is $q_1 1 B q_2$ and $q_2 B R q_1$. Then the TM will erase all the 1s and then stop. If the tape is all 1s, then the TM will never stop.

Although it is not immediately clear, one can also show that by giving a TM the appropriate set of instructions it can perform the task of adding together any two given integers (Boolos and Jeffrey 1989, p. 26 and p. 32). Another instruction set allows the TM to multiply any two integers. Raising to a power, subtraction,

---

[20] A TM is usually defined just as the set of quadruples; but I am keen not to loose sight of the fact that instructions cannot execute themselves — they need hardware and spacetime! For an idea of what the TM's hardware and storage tape might look like, see Penrose's (1989, p. 47) striking picture. For a discussion of whether spacetime possesses sufficient space, time, and matter for a TM, see Barrow and Tipler 1986.

division are also possible. Indeed after becoming familiar with the TM, it becomes reasonable to suppose that the TM can perform any computation that should count as a mechanical computation — and this despite the apparent limitations of having a set-up with a finite set of states and possible operations, a single tape, and an alphabet consisting of only blanks and 1s.

The TM is therefore is a generalisation of an ordinary computer. In fact we can think of the TM as an ordinary computer, so long as we remain conscious of two fundamental differences: (1) the TM never makes a mistake (e.g. no crashing), and (2) TM has access to as much external storage (paper tape) as it needs to execute properly the instructions. This similarity between real computers and the TM allows us to appropriate some ordinary computer terms. So in relation to TMs, I will speak of 'programs', 'software', 'hardware', 'input', 'output' and the like.

To investigate precisely the extent of the computational power of the TM, we make use of the notion of a *partial function*, i.e. a function that maps part or the whole of N^n to N. (Without loss of generality n can and will here taken to be 1.) A partial function f is said to be *computable* or (as I will say here) *Turing computable* just if there is a TM which, when given an arbitrarily chosen n for which f(n) is defined, will eventually (i.e. after some finite number of steps) evaluate f(n). (For n such that $f(n)$ undefined, TM will not halt.) *TCF* will denote the class of Turing computable functions.

We are already in a position to prove something profound: there are partial functions that are not in *TCF*. This is shown as follows. The finite program of a TM consists of a series of lines of code written in some language. By associating a unique integer with each character of the language, we can assign a unique integer to the program and *ipso facto* to the TM as a whole. In this way, every TM acquires a unique number. These numbers are called *Gödel numbers*. (Although the method of coding outlined is highly non-unique, I will stick with one unspecified coding throughout.) This integer/TM correspondence shows that the total number of possible TMs cannot exceed À_0. In fact the number is exactly À_0 since all the functions given by $f(n)=constant$ are clearly Turing computable and they are À_0 in number. But the number of partial functions is the number of mappings of N onto itself, which is $2^{\text{À}_0}$. Thus some (indeed most) partial functions are not Turing computable.

Turing was the first to find an explicit example of a function that is not Turing computable. It concerns the so-called halting problem: is there a Turing computable function that is 1 if the xth Turing machine acting on input y eventually halts and 0 if the xth TM acting on input y never halts? Turing proved the answer is no. His proof employed a diagonalisation argument of the type first used by Cantor in his proof of the nondenumerability of the real numbers. It is surprisingly simple.

Let φ_x denote the partial function characterised by the xth TM (x is a Gödel number). So $\varphi_x(y)$ is the number (if it exists) generated by the xth TM acting on input y. The halting problem can now be put into this form: is there an f TCF such that

$f(x,y)=1$ if $\varphi_x(y)$ eventually halts;

$f(x,y)=0$ if $\varphi_x(y)$ never halts?

Assume there is such an f. Then it can be used to define a new partial function g such that

$g(x)=1$ if $f(x,x)=0$;

$g(x)$ diverges if $f(x,x)=1$.

(Think of a square grid with axes x and y. The diagonal is x,x: hence the term ‘diagonalisation’.) Now g is computable (when $f(x,x)=1$ the TM instruction set will go into an infinite cycle — a pair q_n11q_n, q_nBBq_n will do). Let z be the Gödel number for g. Then by definition of g, $\varphi_z(z)$ convergent $\Leftrightarrow$ $f(z,z)=0$; but by the initial assumption about f, $f(z,z)=0 \Leftrightarrow$ g(z,z) does not halt. Contradiction.

TCF can be characterised in another way. The class of recursive functions (also called ‘partial recursive functions’ though ‘recursive partial functions’ would seem more apt) denoted by R is the smallest class of functions that contains the basic functions of addition and projection, and is closed under substitution, recursion and minimalisation (defined by Boolos and Jeffrey (1989, chapter 7)). Intuitively R is the class of functions that can be built from basic functions and simple manipulations. The alternative characterisation of *TCF* derives from the fact that $TCF=R$ (*ibid.*, chapter 8).

3.8 Computing the Turing uncomputable.

Previously it was shown how a computer (HAL) in a M-H spacetimes could perform a computational supertask. But to shed light on how this new kind of computation connects with Turing computability theory, I will now talk instead in terms of TMs. So HAL is now the name of a TM. We have seen that a TM, when operating in the usual way, defines in function space a boundary between the computable and the uncomputable. One of the aims in this section is to investigate the corresponding boundary or boundaries defined by TMs operating in M-H spacetimes.

To simplify matters in this section, no account will be taken of the physical plausibility of the spacetimes under consideration or of the behaviour of the matter fields they support. Moreover, it will be assumed that every spacetime permits Turing machines of any size to operate unproblematically and that communication to future events is always possible.

Generally speaking a problem is said to be *Turing solvable* if there is a TM that can solve the problem after a finite number of steps. For particular kinds of problems, the words 'computable' or 'decidable' may be used in place of 'solvable'. (Rough guide: 'solvable' is used generally, 'decidable' applies to 'yes-or-no' problems, 'computable' applies to problems with three or more possible answers, e.g. functional problems.) Here I shall say, somewhat informally, that a problem is *solvable in a spacetime* (M, g_{ab}) if there is an observer O M who can initiate a procedure comprised of only Turing machines and ordinary communication devices which will deliver the problem's solution to O after a finite span of O's proper time.

We say that a relation $R(x)$ is *Turing decidable* if the function C_m defined by $C_m(x)=1$ if $R(x)$ holds and $C_m(x)=0$ if $R(x)$ does not hold is Turing computable. C_m is known as the *characteristic function* of R. (R maybe an n-place relation in which case 'x' stands for 'x_1,. . .,x_n'.) Another way to phrase this is to say that $R(x)$ is Turing decidable if there is both a Turing machine TM$_1$ acting on input x (written TM$_1(x)$) that will eventually halt if and only R holds and a Turing machine TM$_2$ so that TM$_2(x)$ that will eventually halt if and only if R does not hold. It follows that if the domain of R is finite, then R is Turing decidable. That is to say, a finite set of yes-no questions is always Turing decidable.

I define *decidable in an M-H spacetime* analogously.

I want now to show that there is a rift between these two notions of decidability. Previously it was shown that the truth value of the Goldbach conjecture and Fermat's last theorem can be decided from (say) the axioms of Peano arithmetic in the M-H spacetime in figure 31, and indeed in any M-H spacetime. Now the truth value of Fermat's last theorem can be decided by a TM operating in the ordinary way using the axioms of Peano arithmetic (we know this because it has been proved true), but whether the Goldbach conjecture is also Turing decidable from these axioms is currently unknown. In any case, these examples will not demonstrate the rift: what is needed is a known Turing undecidable problem that can be shown to be M-H decidable. Here are two examples.

(1) *The halting problem* (Boolos and Jeffery 1989, chapter 3). Rather than treating this as function problem, as we did above, think of it as decision problem: the problem of deciding if an arbitrarily given TM(x) will or will not eventually halt. Working in a M-H spacetime (M, g_{ab}), adopt the following procedure. Let TM(x) move along λ M, having first primed TM(x) to signal to q M if and only if TM(x) halts. The question is settled at q.

(2) *The decision problem for first-order logic* (*ibid.*, chapter 10). This is the problem of deciding whether an arbitrary sentence of first-order logic is valid or invalid. There is a procedure for enumerating all valid sentences of first-order logic, so there is Turing machine, TM_1, that will eventually halt if and only if a given sentence S of first-order logic is valid (*ibid.*, p. 142). However there is no such TM that will eventually halt on input S if and only if the arbitrarily given sentence S is invalid. Thus first order logic is Turing undecidable. Yet it is decidable in a M-H spacetime. The procedure is obvious: let TM_1 move along λ M, having first primed TM_1 to signal to q if and only if TM_1 halts. Upshot: a signal at q means the sentence is valid, no signal at q means the sentence is invalid.

We saw that in order to decide a relation R two particular Turing machines, TM_1 and TM_2, were required. If however only one (or both) of that pair TM_1, TM_2 exists, then R is said to be *partially Turing decidable*. Problems (1) and (2) above are clearly of this kind. It is now evident that:

Proposition 3.8.1. If a problem P is partially Turing decidable then P is decidable in a Malament-Hogarth spacetime.[21]

There is 'set' version of this result. Let A be a subset of N. If the function f defined by $f(x)=1$ if x A and $f(x)=0$ if x A is in *TCF*, then we say that A is a *recursive set*. (f is called the *characteristic function* of A.) On the hand, if the function f given by $f(x)=1$ if x A, f is defined or undefined if x A is a *TCF*, then we say A is *recursively enumerable*. The idea is that one can enumerate both the members and the non-members of a recursive set, whereas only the members of a recursively enumerable set can be necessarily enumerated. The set version of proposition 3.8.1 is now clear.

---

[21] I will use the term 'proposition' with regard to results about M-H computability, though I recognise that these are not results of pure mathematics.

Proposition 3.8.2. Membership of a recursively enumerable set A is decidable in any Malament-Hogarth spacetime. Also, membership of a set B whose complement is a recursively enumerable set is decidable in any Malament-Hogarth spacetime.

(Hereafter I assume that the reader is familiar with the various modes of speech used to express solvability. For example if A is a set then ‘the characteristic function of A is Turing computable’, ‘membership / non-membership of A is Turing decidable’, ‘A is Turing decidable’, are all ways of saying the same thing.)

Now any recursively enumerable set can written in the form $\{x \mid \quad y\, R(x,y)\}$, where R is a recursive relation (Rogers 1987, p. 66). So by proposition 3.9.2 any set A of the form $A=\{x \mid \quad y\, R(x,y)\}$ is M-H decidable. This raises a question: are sets of a more complicated form, say, $B=\{x \mid \quad y \quad z\; S(x,y,z)\}$, where S is recursive, necessarily decidable in a M-H spacetime?

The answer is not clear. Certainly naive attempts fail. For example, to decide if a number u is in B using the machine in figure 31 — what I will call, for reasons given later, a *SAD*$_1$ *machine* — one could set y=1 and run through the zs to check if $\quad z\; S(u,1,z)$ holds; if it does not then one would need to check $\quad z\; S(u,2,z)$; and so on. The reason why this method fails of course is that the λ-curve is used up on the first z-sweep.

If the SAD$_1$ machine cannot cope with double quantifiers, then it seems likely it could not decide arithmetic. On the other hand perhaps any deficiencies of the SAD$_1$ machine could be overcome by utilising more complicated hardware configurations. These ideas lead us into the next section.

3.9 Deciding arithmetic

‘Reeling and writhing, of course, to begin with’, the Mock Turtle replied; ‘And then the different branches of Arithmetic — Ambition, Distraction, Uglification, and Derision’.

LEWIS CARROLL, *Alice in Wonderland*

First let us make sure we know what we are talking about. The language of arithmetic, nick-named L, consists of the symbols 0, 1, 2, ..., +, ×, =, together with the logical notions: ¬ (not), (and), (or), → (implies), (universal quantifier), (existential quantifier). L also consists of variables e.g. $x, y, z, \ldots$ and brackets (and). The sentences of L are the meaningful finite strings of symbols taken from L which have no free variables; e.g. 2+3=5. Sentences of L are often called arithmetical sentences, but here they will often be referred to just as sentences.

Sentences acquire truth value through an *interpretation*, and the interpretation to be adopted here will be the *standard* one (see Boolos and Jeffrey 1989, p. 199). Thus sentences are ‘true’ or ‘false’ in accordance with the ordinary use of those labels. So

$$2+3=7,$$

$$\neg \{\ w\ \ x\ \ y\ \ z\,[(x+1)^{w+3} + (y+1)^{w+3} = (z+1)^{w+3}]\}$$

are two sentences, of which the first is false and the second (a statement of Fermat’s Last Theorem) is true.

Now the key result *about* arithmetic is due Gödel ((Boolos and Jeffrey 1989, chapter 15). The version of his incompleteness theorem required here is this: there is no TM that will halt (proof ended) if an arbitrarily given sentence is true, and there is no TM that will halt if an arbitrarily chosen sentence is false.[22] This is result (i) below. It is followed by four other standard results that we need in order to proceed.

Standard results 3.9.1.

(i) Arithmetic is not partially Turing decidable (Boolos and Jeffrey 1989, chapter 15).

(ii) Quantifier-free arithmetic sentences are Turing decidable (just calculate the answer in the obvious way).

(iii) There is a Turing machine that will eventually translate an arbitrary sentence S into a logically equivalent sentence, S', where quantifiers of S' occur at the extreme left (*ibid.*, p. 109).

(iv) There is a TM that will eventually translate an arbitrary sentence S containing two juxtaposed quantifiers of the same type into a logically equivalent sentence, S', with one quantifier of that type in place of the previous two (*ibid.*, p. 109).

---

[22] Another way to phrase this is to say that the true (or false) sentences of arithmetic do not form a recursively enumerable set.

Sentences that have undergone the transformation described by (iii) are said to be in *prenex* form; those transformed according to (iv) are said to be in *normal* form. I will assume hereafter that all sentences are in prenex normal form.

Gödel showed that sentences have relational counterparts (Rogers 1987, § 14.4). To make this precise first requires a definition. An n-place relation R is said be in the *arithmetical hierarchy* if and only if it is recursive or, for some $m \geq 1$, can be expressed as

$$\{<x_1,\ldots,x_n> \mid (Q_1 y_1)\ldots(Q_m y_m) S(x_1,\ldots,x_n,y_1,\ldots,y_m)\},$$

where Q_i is either or , and S is an $(n+m)$-place recursive relation.

(v) There is a TM which when given an arbitrary sentence S will eventually construct an n-place relation R in the arithmetical hierarchy such that

S true in the standard interpretation if and only if $(Q_1 x_1)\ldots(Q_n x_n) R(x_1,\ldots x_n)$ holds,

where successive Q_is alternate between and .

R is called the *relational counterpart* of S. If the relational counterpart of a sentence S has $Q_1 =$ and is n-place, then we call S a Σ_n-*sentence*; if the relational counterpart is n-place and $Q_1 =$ we call S a Π_n-*sentence*.

In fact the converse to (v) is also true: to any relation in the arithmetic hierarchy there corresponds a sentence. This correspondence means we can legitimately use the word ‘sentence’ to refer to a sentence and its relational counterpart. This I will do often.

Now it follows from (ii) that arbitrary sentences (=relational counterparts) of either the form $nS(n)$ or $nS(n)$, where S is recursive, are partially Turing decidable. So by proposition 3.8.1, they are both decidable in any M-H spacetime. This fact, together with (iv) above, implies the following:

Proposition 3.9.2. Arbitrary purely existential or purely universal sentences are decidable in any Malament-Hogarth spacetime.

(Incidentally, the Goldbach conjecture and Fermat’s last theorem can both be stated as purely universal sentences.)

But because of (i), proposition 3.8.1 cannot be used to show that arithmetic is decidable in any M-H spacetime. What I want to show now is how in more subtle kinds of M-H spacetimes TMs can decide sentences of quantificational orders greater than 1. Then later in section 3.12 I will address the question of whether it is truly impossible for, say, an arbitrary double-quantifier sentence to be decided using the SAD_1 machine.

Definition 3.9.3 (Hogarth 1994). In a spacetime (M, g_{ab}), a set of non-intersecting open regions O_i, i=1,2,... such that (1) for all i O_i $I^-(O_{i+1})$ and (2) there is point q M such that for all i O_i $I^-(q)$, is said to form a *past temporal string* or just *string*.

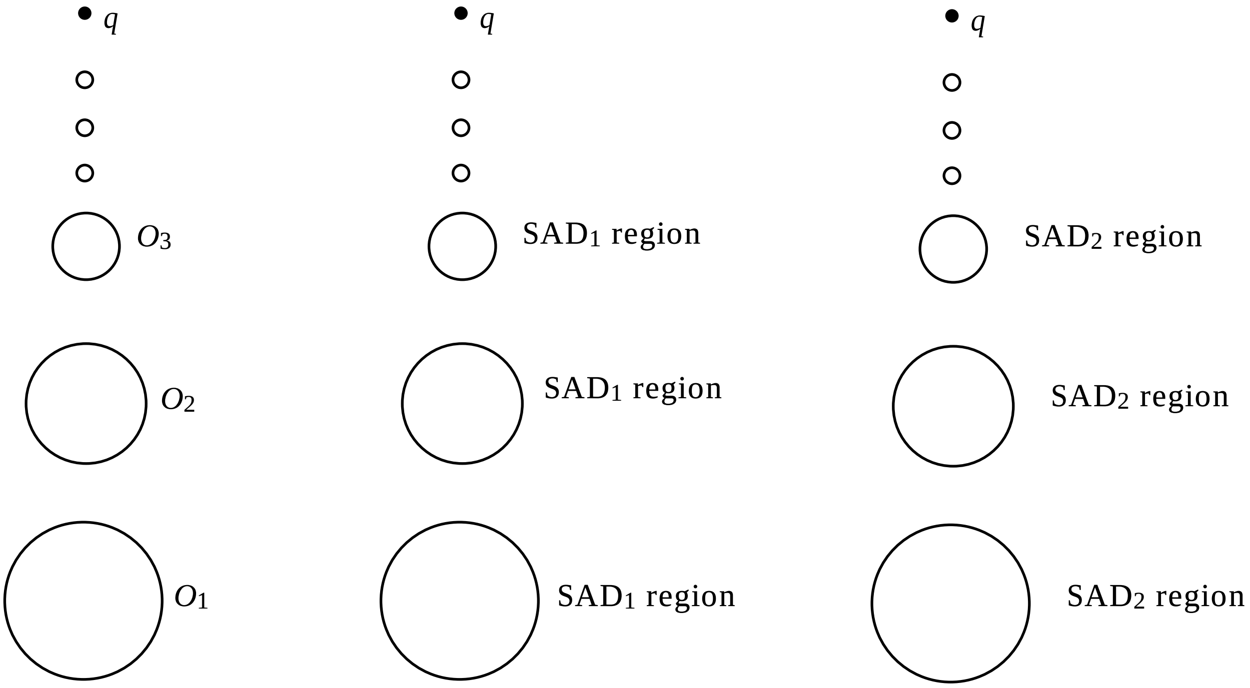
See figure 39(i).

Definition 3.9.4 (Hogarth 1994). A spacetime (M, g_{ab}) is an *nth-order arithmetical sentence deciding* (denoted SAD_n) spacetime if the n conditions contained in the following scheme are satisfied.

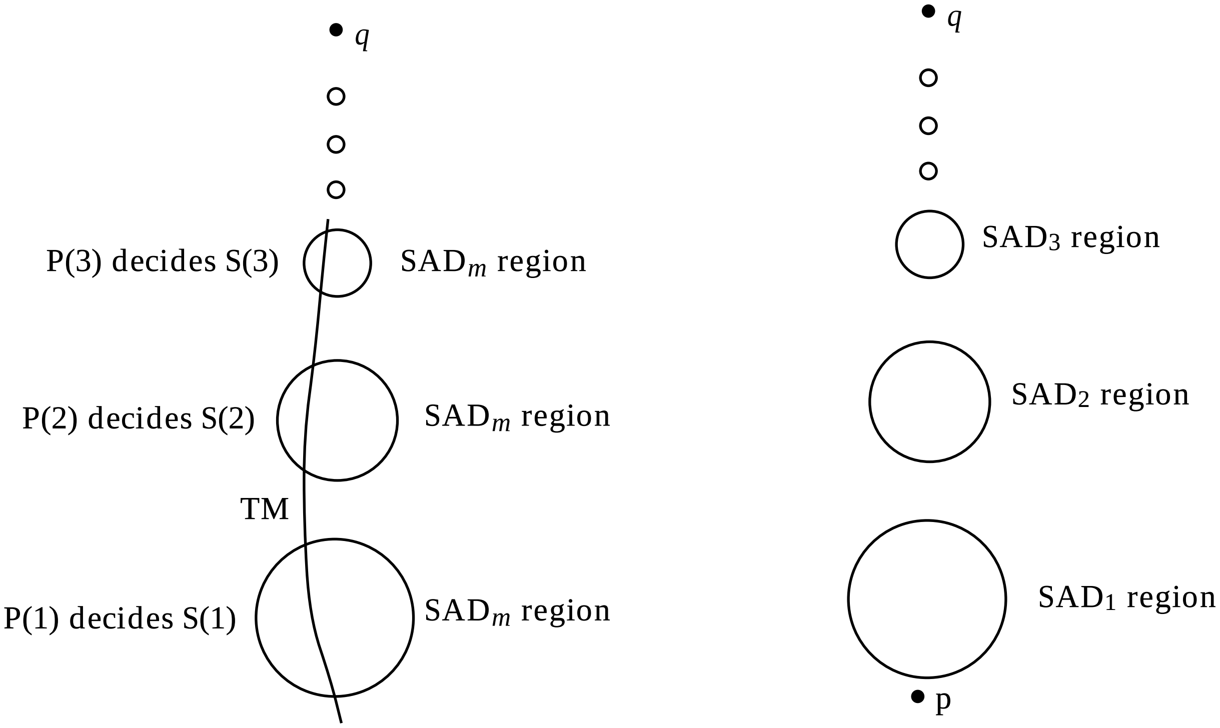
If n=1, (M, g_{ab}) is a M-H spacetime.

If n>1, (M, g_{ab}) admits a string of SAD_{n-1} spacetimes.

This is simpler than it might first appear. According to definition 3.9.4, a SAD_1 spacetime is a M-H spacetime, a SAD_2 spacetime is a spacetime that contains a string of SAD_1 spacetimes (figure 39(ii)), a SAD_3 spacetime is a spacetime that contains a string of SAD_2 spacetimes (figure 39 (iii)), and so on.



(i) A past temporal string. (ii) A SAD$_2$ spacetime. (iii) A SAD$_3$ spacetime.



(iv) Diagram used in lemma 3.9.5. (v) An AD spacetime.

Figure 39.

The efficacy of SAD spacetimes to decide sentences in arithmetic derives essentially from the following result.

Lemma 3.9.5. Suppose $S(1)$, $S(2)$, $S(3)$,... is a sequence of sentences that can be enumerated by a TM. Suppose further that for each $m \geq 1$, $S(m)$ is decidable in any SAD$_n$ spacetime. Then $mS(m)$ and $mS(m)$ are both decidable in a SAD$_{n+1}$ spacetime.

Proof. This consists of showing how appropriately chosen and appropriately located hardware can be used to decide $mS(m)$ and $mS(m)$. Part of this involves a TM travelling along a λ-curve that picks up one hour of proper time in each SAD$_\mathrm{m}$ component (this is possible because every such component admits a λ-curve), as depicted in figure 39 (iv). TM is primed to generate $S(1)$ in the lead up to the first component and to signal that sentence to the first component. Similarly TM is also primed to generate, for $m>1$, $S(m)$ in the $(m–1)$th component and to signal that sentence to the mth component. In this way, the mth component receives the sentence $S(m)$.

Now let $P(m)$ denote the procedure that by hypothesis decides $S(m)$. The procedure for deciding $mS(m)$ consists of adding to each $P(m)$ transmitting devices and receivers which operate as follows.

> $P(1)$ signals to q and $P(2)$ if and only if $S(1)$ holds.
>
> For $m>1$, $P(m)$ signals to $P(m+1)$ if and only if $P(m–1)$ signals to $P(m)$.
>
> For $m>1$, $P(m)$ signals to q and $P(m+1)$ if and only if $S(m)$ holds and $P(m)$ has not received a signal from $P(m–1)$.

This procedure ensures that a *single* signal is sent to q if and only if there is an m such that $S(m)$ holds. (The signal is actually sent by $P(x)$, where x is the smallest integer for which $S(x)$ holds.) Upshot: a signal at q means $mS(m)$ holds (true), no signal at q means $mS(m)$ does not hold (false).

The procedure for deciding $mS(m)$, given below, is similar except this time a single signal at q means $mS(m)$ does not hold, no signal at q means $mS(m)$ does hold.

> $P(1)$ signals to q and $P(2)$ if and only if $\neg S(1)$ holds.
>
> For $m>1$, $P(m)$ signals to $P(m+1)$ if and only if $P(m–1)$ signals to $P(m)$.

For $m>1$, $P(m)$ signals to q and $P(m+1)$ if and only if $\neg S(m)$ holds and $P(m)$ has not received a signal from $P(m-1)$.

Thus $mS(m)$ and $mS(m)$ are seen to be decidable in (M, g_{ab}). ♦

We have seen already how single-quantifier sentences can be decided in SAD_1 (=M-H) spacetimes. Now suppose that we are given an arbitrary double-quantifier sentence $S=Q_1(y)Q_2(z)F(y,z)$, where F is a recursive relation and Q_i is either or . Writing S as $Q_1(y)\{Q_2(z)F(y,z)\}$ means that lemma 3.9.5 can be applied (F recursive obviously implies the sentences $Q_2(z)F(1,z)$, $Q_2(z)F(2,z)$,... can be listed in order) to show that double-quantifier sentences are decidable in SAD_2 spacetimes. A further application of the lemma shows that triple-quantifier sentences can be decided in SAD_3 spacetimes. And continuing in this way we arrive at the following fundamental result.

Proposition 3.9.6 (Hogarth 1994). An arbitrary n-tuple quantifier sentence can be decided in a SAD_n spacetime.

In fact, these different order SAD spacetimes can be accommodated within a single spacetime.

Definition 3.9.7 (Hogarth 1994). A spacetime (M, g_{ab}) is an *arithmetic deciding* (AD) spacetime just when (M, g_{ab}) admits a string with open regions O_1, O_2, O_3,... such that for each $n\geq1$, (O_n, g_{ab}) is a SAD_n spacetime.

In figure 39(v), an observer at p can decide an arbitrary sentence S by communicating it to the SAD_n region that decides sentences of that order. Thus we have the following:

Proposition 3.9.8 (Hogarth 1994). Arithmetic is decidable in an AD spacetime.

3.10 Assembling the hardware

If the appropriate hardware is in place then arithmetic will succumb. But it is natural to wonder how the hardware might be installed. (Not that theorists worry much about this with regard to the TM!) What follows is a *prima facie* reasonable method of performing that task; see figure 40.

The various paths of the hardware through spacetime are represented by worldlines, and the idea of the method is to begin with one worldline at the initial event p and to have a process of worldline branching that results in each SAD_1

component being populated by a λ-curve and every other component of every string being populated by at least one worldline. (Recall that the SAD_1 spacetimes accommodate TMs travelling on λ−curves, while all the other components accommodate communicating devices à la lemma 3.9.5.)

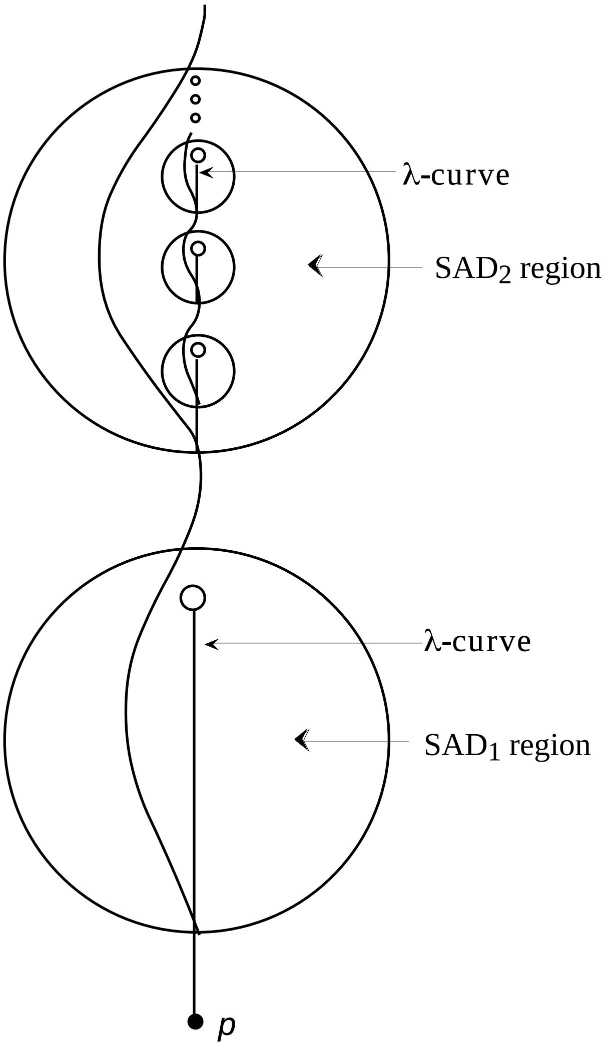


Figure 40. The 'hardware tree' grows up into the AD spacetime.

Conditions for hardware construction 3.10.1

(1) A worldline that meets a SAD_n component, $n \geq 1$, must bifurcate, with one worldline branch extending to the next component of that string and the other branch entering the component and extending to the first component of the string on the 'next level down'.

(2) Two kinds of worldline must follow the first available λ-curve: the one that enters the first SAD_1 component and any one that is constrained by (1) to enter a SAD_1 component.

Figure 40 illustrates the process at work on the SAD_1 and SAD_2 stages of the AD spacetime. The fact that the hardware is constructed by a process that resembles actual tree growth adds a certain realism to the method.

But there is one potential problem that needs to be addressed: might there be an infinite amount of hardware matter residing in some compact set? If yes then the situation is hopelessly unphysical because there would be at least one point where the density of matter was infinite. In fact the worry is not that (even one) entire TM could reside in a compact set, for that would entail an imprisoned future endless curve (λ) — which according to lemma 3.5.2 is impossible. Rather the worry is the possibility of an infinite number of finite TMs intersecting a compact set, as in C_1 in figure 41.

To guard against this requires an additional condition.

> (3) Bifurcations points must occur further and further along the λ-curves with each new component.

I grant that this is somewhat vague but the idea is clear enough. The method now involves moving the bifurcation points further and further along the λ-curves, so that eventually the hardware is pushed outside any given compact set, e.g. C_2 as depicted in figure 41. C_2 is therefore intersected by only a finite number of finite parts of TMs — and the world is safe from overbalancing.

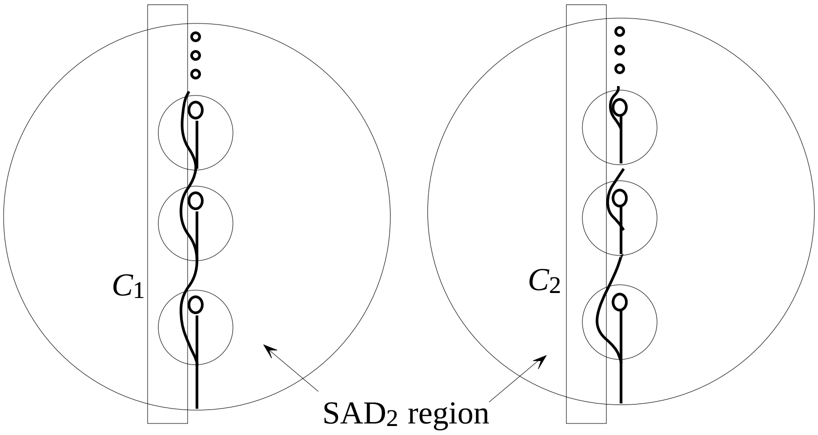


Figure 41. The diagram on the left illustrates the method of AD assembly according to (1) and (2). The compact region C_1 could conceivably contain an infinite amount of hardware matter. On the other hand, by adding condition (3) it seems that the corresponding C_2 could possess only a amount finite matter.

(Admittedly there remains the worry that an infinite amount of mass might reside in a region which is non-compact but of finite volume. I am not sure whether or not this can happen. Are there any SAD_1 spacetimes with finite volume? In any case, the two examples in the next section suffer no such pathology.)

Now if the instructions (1), (2), (3) that serve to build the hardware are supplemented by the software instructions required for deciding arithmetic à la lemma 3.9.5, then the result is a machine that can decide arithmetic. The machine (hardware and software) can therefore be assembled from a *finite* plan. This is one feature it shares with the TM.

3.11 Two examples of AD spacetimes

The first example is really a toy.[23] Start with Minkowski spacetime (^{4}, η_{ab}) and choose a compact set C ^{4}. Now draw a closed inertial line segment ν C. About this ν, define regions O_1, O_2, O_3,… with inclusion relations appropriate for all the strings of an AD spacetime, in such a way that ν intersects every component, as depicted in figure 42. Then choose a scalar field Ω on M such that Ω=1 outside C and Ω tends rapidly to infinity as the line ν is approached. Remove ν. Then (^{4}–ν, $\Omega^2\eta_{ab}$) is an AD spacetime because O_1 is a SAD_1 spacetime, O_2 is a SAD_2 spacetime, and so on. Moreover, every component of every string has infinite volume (despite appearances!), thus giving the TMs the space they need.

Although the corresponding AD machine consists of an infinite number of infinitely large regions, each with its own communication devices and TMs, it can still be contained in a box, e.g. the one depicted in figure 42, with *finite* spatio-temporal surface area (recall the Tardis). So in this regard this the AD machine is no different to an ordinary desktop computer.

In Hogarth 1994 I conjectured that anti-de Sitter spacetime (in its universal covering form) is AD. This can now be proved.

---

[23] When I showed my first example of an AD spacetime to John Norton he rightly accused me of having created a monster. It involved first removing a complicated countably infinite set of points that fell on a line segment, then arranging a conformal factor that blew-up only at each such point, and finally removing the points. Norton suggested that I should simply use a conformal factor that blew-up on the whole line segment and then remove that line. I followed his advice.

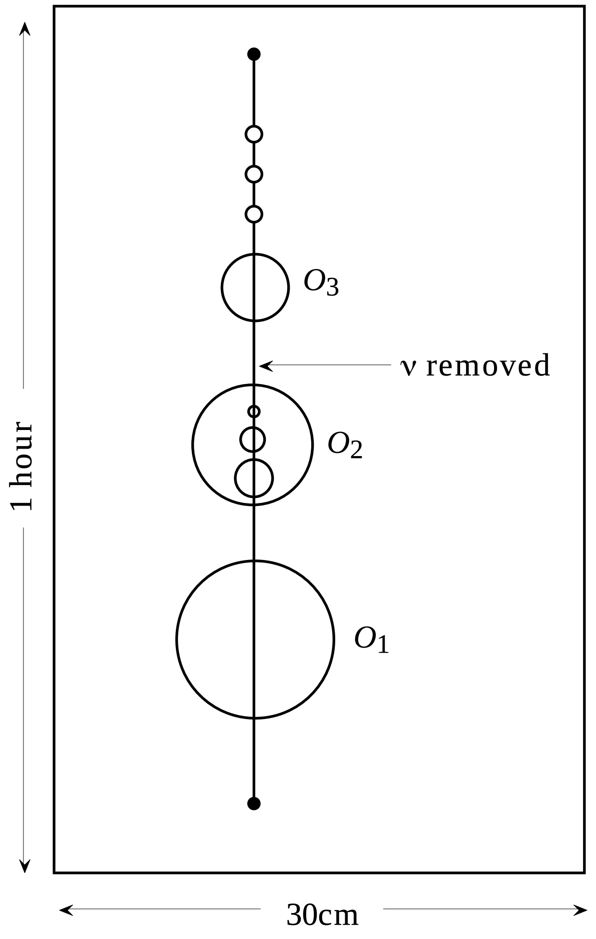


Figure 42. A toy AD spacetime.

Proposition 3.11.1. The universal covering form of anti-de Sitter spacetime is AD.

Proof. I will work with a 2-dimensional spacetime, but the argument readily extends to four dimensions. As we saw in section 3.3, the 2-dimensional line element takes the form

$$ds^2 = \cosh^2 r \, dt^2 - dr^2$$

The key result is this: if $\pi/2 < a < b < \pi$, then the region of spacetime given by $\{(t,r) \mid a < t < b, 0 \leq r < \infty\}$ is M-H.

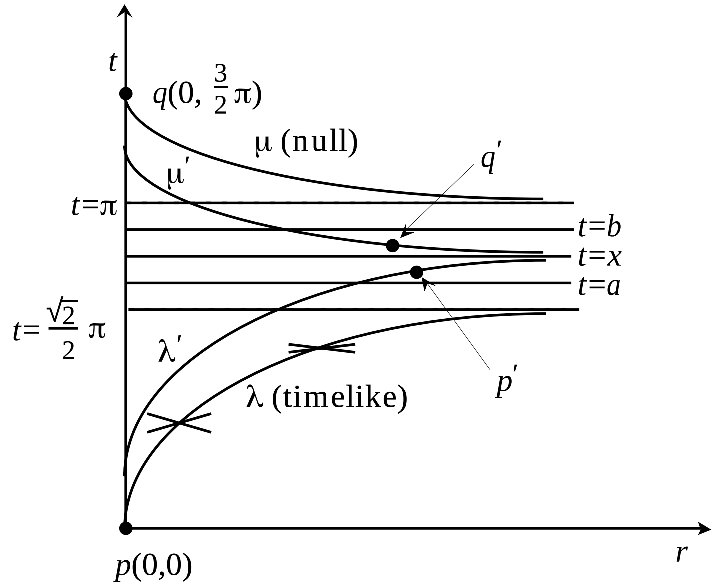


Figure 43. Any temporal sandwich of anti-de Sitter spacetime is Malament-Hogarth.

Choose λ as in the proof, given in section 3.3, that anti-de Sitter M-H: so the length of λ is infinite and it asymptotically approaches the line $t=\frac{\sqrt{2}}{2}\pi$. See figures 33 and 43. It was also shown in section 3.3 that the past null geodesic through q, labelled μ in figure 43, asymptotically approaches the line $t=\pi$. Now notice that since the (2-dimensional but also 4-dimensional) metric does not depend explicitly on the t-coordinate, the metrical structure at (t,r) is identical to that at $(t+constant, r)$. Consequently light cones and timelike curves can be translated along any line $r=constant$. Choose x such that $a<x<b$, and translate λ along $r=0$ into λ', the timelike curve that asymptotically approaches $t=x$, as depicted in figure 43. Likewise translate μ along $r=0$ into μ', the past light cone which asymptotically approaches $t=x$. Now choose a point $p'\in\lambda'$ such that the t-coordinate of p' lies between a and x; and choose a point q' on μ' such that the t-coordinate of q' lies between x and b. It is clear that the part of λ' from p′ onwards is infinitely long and that it lies to the past of q'. Since p', q', and λ' all lie within the region $\{(t,r)\,|\,a<t<b,\ 0\leq r<\infty\}$, this region is therefore M-H.

Now all that remains is to construct open sets of the spacetime with inclusion relations appropriate for all the strings of the AD spacetime. The method is analogous to that used for the previous example and is illustrated in figure 44. ♦

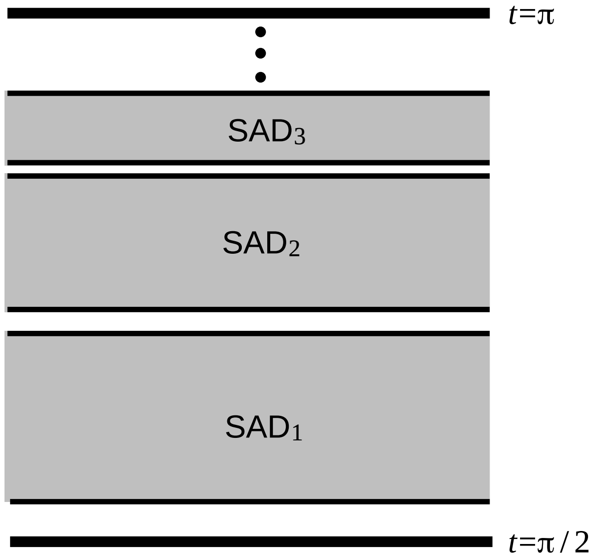


Figure 44. Constructing AD open sets in a temporal sandwich of anti-de Sitter spacetime.

3.12 Pure Computability

Axiomatising the TM-based computers

Textbook Turing computability theory usually begins with an argument aimed at showing that any finite number of ordinary computing machines in any configuration can always be mimicked by a single appropriately programmed TM. The theory then proceeds in terms of this one abstract and easily characterised machine, and thereby manages to transcend irrelevant hardware details. Indeed it is this very process of abstraction that makes Turing computability a branch of pure mathematics.

Now I want to apply this idea to the TM-based computers encountered so far. This will include the TM that operates in the ordinary (à la Turing) way and also a device like a TM except that it exists for only a finite time. I will call the first computer an *ordinary TM* (abbreviation: OTM) and the second a *finite TM* (abbreviation: FTM). FTM_i will denote a FTM that can perform at most i operations. A SAD_n *machine* (abbreviation: SAD_n) is the nth order device that operates (in a SAD_n spacetime) according to the rules in section 3.10. An *AD machine* (abbreviation: AD) is defined similarly. The term *naked Turing machine* (abbreviation: NTM) will be used to refer generally to any TM-based computer that depends on the M-H property. The 'axiomatization' of a particular physical computer is achieved by providing a simple yet computationally equivalent schematic representation — a blue print, if you like. These are depicted in figure 45.

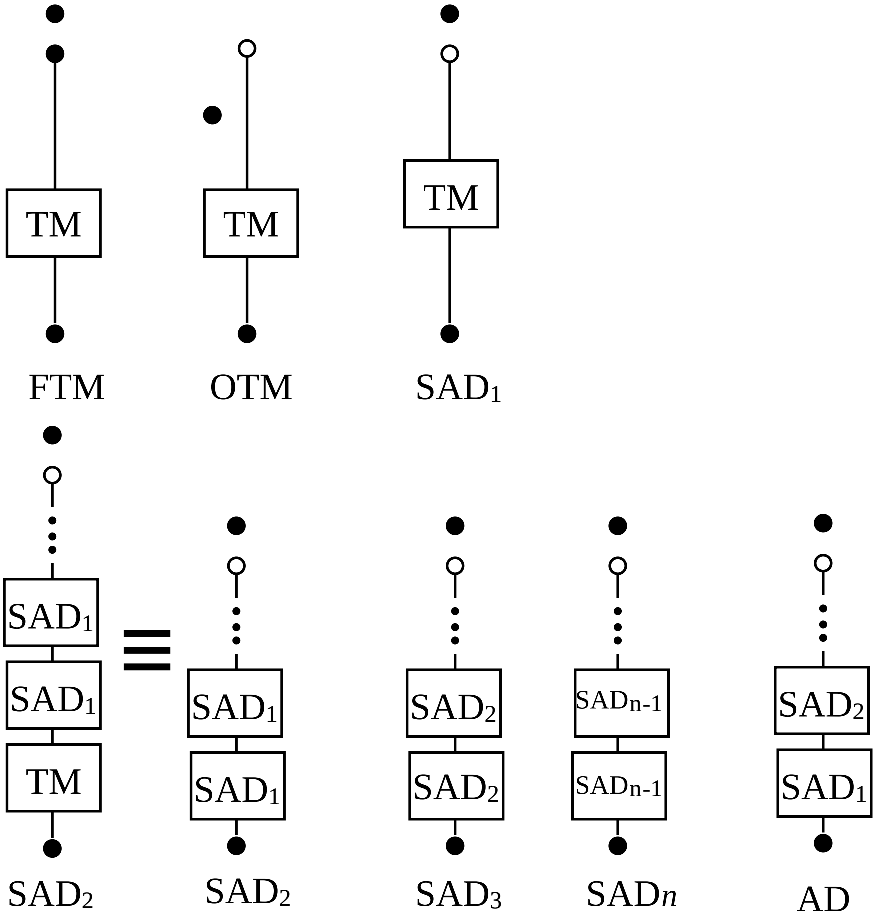


Figure 45. Blue prints of TM-based machines. '≡' means 'is equivalent to'.

The diagrams are based on spacetime diagrams, but the reader can — and from a pure mathematical view point *should* — view them as purely schematic. A filled dot represents a M-H event — now to be called, in the spirit of purity, a *solution event*; a line is a hardware worldline; an empty dot is 'the edge of spacetime at infinity'; three dots means the sequence of hardware boxes continues indefinitely. In the FTM diagram, there is single finite TM that travels some finite temporal distance (time's up), and lies below (in the past) of a particular solution event. Next is the OTM: one TM travels to infinity and the solution event lies ahead of a finite part of the TM. Next are the NTMs. The SAD$_1$ has one TM travelling to infinity, but this time the whole of TM's worldline lies below the solution event. SAD$_2$ is a string to infinity of SAD$_1$ computers: each box houses a SAD$_1$ computer. Two representations are given: the one the left has a TM at the beginning to remind us where the software that drives each of the SAD$_1$s resides. The representation on the right is used when no such reminder is needed. Last in the figure is the AD machine: a string of increasingly powerful SAD machines.

To reflect the physical NTMs, three self-explanatory rules must govern the 'transfer of information' within the schemata.

Conditions 3.12.1.

(1) a lower box can signal to higher box or to the solution event;

(2) each box and the signal event can receive at most one signal (no swamping);

(3) downward signalling is forbidden.

The power of the NTMs

In the previous section we showed how arbitrary $\sum_n$-sentences or Π_n-sentences could be decided in an M-H spacetime using a specific hardware set-up (now called a SAD$_n$). This result carries over naturally to the pure case, and so we have:

Proposition 3.12.2. A SAD$_n$ can decide an arbitrary $\sum_n$-sentence (or Π_n-sentence).

Proposition 3.12.3. An AD can decide arithmetic.

But the AD machine is not omniscient, as the corollary of the following result shows.

Proposition 3.12.4. There are exactly À$_0$ functions which the AD machine can compute.

Proof. Since each AD machine is characterised by the finite programme that drives it, the set of AD machines can be given a Gödel numbering. The proof now follows the proof of the corresponding OTM result (section 3.7). ♦

Corollary. There exist functions which the AD machine cannot compute.

Proposition 3.12.2 is a positive result about the power of SAD$_n$ machine, but it raises a question: just how far does the power of the SAD$_n$ machine extend? The case of n=1 has been investigated by Earman and Norton (1994), who found that the SAD$_1$ machine cannot decide arbitrary sentences with two quantifiers. I will now show that this statement generalises to any n: SAD$_n$ cannot decide arbitrary sentences with n+1 quantifiers. Unfortunately the proof is not a trivial generalisation of the n=1 case, so I will first need to rehearse some more standard theory. (For a more detailed account see Rogers 1987, chapter 14.)

Let $\sum_n$ (respectively Π_n) denote the set relations expressible by a series of n quantifiers that begin with (respectively) and act on a recursive relation. For example, if the two-place relation $R(u,v)= \quad x \quad yS(x,y,u,v)$, where S is recursive, then $R \quad \sum_2$. Let Δ_n denote the set of relations that are expressible in both such forms; hence $\Delta_n=\sum_n \cap \Pi_n$.

These three results are standard: $\sum_n \quad \Pi_n \quad \sum_{n+1} \cap \Pi_{n+1}=\Delta_{n+1}$; $(\sum_n - \Pi_n) \neq$; $(\Pi_n - \sum_n) \neq$. It follows that the classes $\sum_0$, $\sum_1$, $\sum_2$, $\sum_3$,... and the classes Π_1, Π_2, Π_3,... each form a strictly increasing sequence. This result is known the *arithmetical hierarchy theorem*.[24] It can interpreted as saying that mathematics becomes genuinely more complex as the number of (alternating) quantifiers increases.

The notion of an *oracle* was invented by Turing. One can think of it as a 'black box' that possesses by fiat certain well-defined powers of computation (e.g. it can decide only recursively enumerable sets). The corresponding *oracle machine* is a TM that can consult the oracle during any step of its operations. So, for example, by an Π_1-*oracle-machine* I mean a OTM whose repertoire of ordinary operations is supplemented by the capacity to receive, when required, correct answers to any question of just the form 'does $R(x)$ hold?', where $R \quad \Pi_1$.

Oracles and the arithmetical hierarchy come together in this version of *Post's theorem*. A Π_n-oracle machine ($\equiv\Sigma_n$-oracle machine) can decide relations in precisely Δ_{n+1}. This shows that if one had complete 'access' to a particular level of the arithmetical hierarchy, then one could recursively access no more than a small subset of the next level up.

I will, where appropriate, adopt the standard practise of abbreviating relations like $(\quad x)(\quad y)(\quad z)R(u, x,y,z)$ & $(\quad x)S(u, x)$, where R and S are recursive, to & ; that is, I will indicate only the quantifier symbols and logical connectives. $_n$ (respectively, $_n$) will denote an alternating string of n-quantifiers beginning with (respectively,).

Using the *Tarski-Kuratowski algorithm* any such relation expression involving quantifiers and logical connectives can be reduced to an equivalent series of alternating quantifiers (which in turn allows one to establish where a given relation is located in the arithmetical hierarchy). The details of the algorithm are simple but too long for inclusion here (*ibid.*, p. 309). In any case the method is quite intuitive. Here are three examples. ('≡' means 'is equivalent to'.)

$$\neg \quad \equiv$$

---

[24] The special case '$\sum_0 \neq \sum_1$' represents another statement of Gödel's incompleteness theorem.

$\neg\ (\quad \&\quad) \equiv \neg\ (\qquad) \equiv \neg \quad \neg \quad \equiv$

$[\quad \Leftrightarrow\ \neg\quad] \equiv \quad [[\qquad \neg\ ]\ \&[\neg \qquad]] \equiv \quad [[\qquad]\ \&[\qquad]] \equiv \quad [\quad \& \quad] \equiv$
$\equiv$

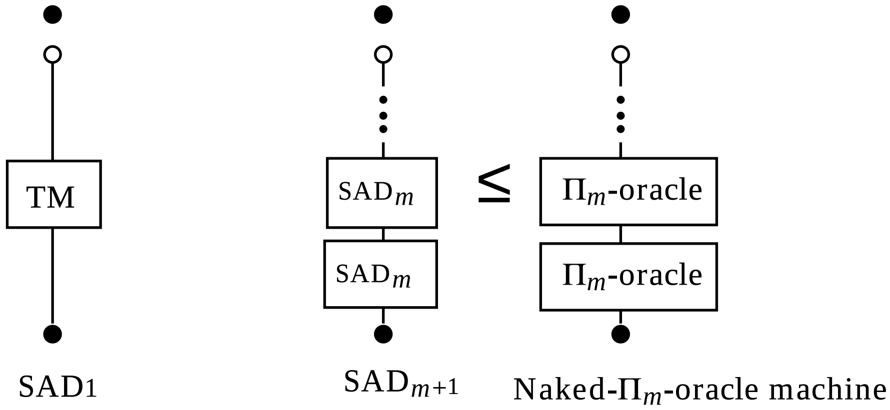


Figure 46. Diagram used in the proof of lemma 3.12.5.

We are now in position to state the main result.

Proposition 3.12.4 $\mathrm{SAD_n}$ cannot decide an arbitrary Π_{n+1} sentence (or an arbitrary Σ_{n+1} sentence).

Since the proof is rather long, it is helpful first to prove the following:

Lemma 3.12.5. Suppose $S(z)$ is 1-place relation that is SAD_n decidable. Then $S(z)$ is of the form $S_1(z)\ \ S_2(z)$, where $S_1(z)\ \ \Pi_n$ and $S_2(z)\ \ \Sigma_n$.

The proof is by induction.

The case n=1. Keep an eye on the SAD_1 schematic in figure 46. If the SAD_1 is to decide $S(z)$ then it must perform in one of four ways:

(1) for all z, there is a signal (to arrive at the solution event);

(2) for all z, no signal means that $S(z)$ does not hold;

(3) for all z, no signal means that $S(z)$ holds;

(4) for some z no signal means $S(z)$ does not hold, and for some z no signal means $S(z)$ holds.

(In case (4) it must be a Turing computable matter which z means ‘signal implies does not hold’ and which z means the opposite.)

Case (1) implies that the TM program must halt for each z, so $S(z)$ must take the form $S(z)=R_a(z)$, where R_a is a recursive relation. Thus $S(z)$ is in Π_0.

Case (2) implies that when $S(z)$ holds, the TM halts after a finite number of steps. This means that for each z a number is reached which halts the TM program. Thus $S(z)$ is expressible as $S(z)=\ yR_b(y,z)$ for some recursive relation R_b; i.e. $S(z)$ is .

Case (3) implies that the TM will halt only if S(z) does not holds, that is, when $\neg S(z)$ does hold. From case (2), this implies that $\neg S(z)$ is ,i.e. S(z) is .

Case (4) implies that for some z $S(z)$ is as in case (2), so S is ; and for some z, $S(z)$ is as in case (3), so $S(z)$ is . Thus $S(z)$ is .

Collecting together cases (1) through (4) proves the case n=1.

The general case. Now suppose the lemma is true for some arbitrary but fixed m, i.e. if $S(z)$ is SAD$_m$ decidable, then $S(z)$ is a (one-place) relation of the form $S_1(z)\ \ S_2(z)$, where $S_1(z)\ \ \Pi_m$ and $S_2(z)\ \ \Sigma_m$.

Consider the SAD$_{m+1}$ schematic in figure 46. It consists of a string of SAD$_m$ components. Relations that a SAD$_m$ component can decide can, by the hypothesis of the previous paragraph, be decided by a Π_m-oracle. Thus by replacing each SAD$_m$ component by a Π_m-oracle, there is no loss of computational power. The question now is: what is the power of this 'naked-Π_m-oracle machine'?

Suppose $T(z)$ is a 1-place relation that is decidable by the naked-Π_m-oracle machine. Then the machine must perform in one of four ways:

(1′) for all z, there is signal;

(2′) for all z, no signal means that $T(z)$ does not hold;

(3′) for all z, no signal means that $T(z)$ does hold;

(4′) for some z no signal means $T(z)$ does not hold, and for some z no signal means $T(z)$ does hold.

By analogy with n=1 case and applying Post's theorem, we see that:

Case (1′) implies that $T(z)\ \ \Delta_{m+1}$.

Case (2′) implies that $T(z)=\ yR_b(y,z)$, where $R_b(y,z)\ \ \Delta_{m+1}$. This implies that $T(z)$ is [$_{m+1}$] and $T(z)$ is [$_{m+1}$], so by the Tarski-Kuratowski algorithm $T(z)$ is $_{m+1}$, Case (3′) implies that $T(z)$ is $[\ _{m+1}]$ and $T(z)$ is $[\ _{m+1}]$, so by the Tarski-Kuratowski algorithm $T(z)$ is $_{\mathrm{m+1}}$.

Case (4′) implies that $T(z)$ is $_{m+1}\quad _{m+1}$.

Collecting together cases (1′) through (4′) shows that $T(z)$ naked-Π_m-oracle machine decidable implies $T(z)$ is $_{m+1}\quad _{m+1}$. But we know that $T(z)$ SAD$_{m+1}$ decidable implies $T(z)$ naked-Π_m-oracle machine decidable. Thus $T(z)$ SAD$_{m+1}$ decidable implies $T(z)$ is $_{m+1}\quad _{m+1}$.

Hence the mth case implies the $(m+1)$th case. Since the $m=1$ case is true, the proof is complete.♦

Corollary. $S(z)\quad \Delta_{n+1}$.

Proof. Apply the Tarski-Kuratowski algorithm. ♦

Proof of proposition 3.12.4. By the Kleene hierarchy theorem there are relations in Π_{n+1} that are not in Δ_{n+1}. Thus if $U(z)$ is one such 1-place relation then the corresponding family of sentences cannot be decidable by SAD$_n$. The same applies to Σ_{n+1}.♦

Combining propositions 3.12.2 and 3.12.4 gives:

Proposition 3.12.5. The SAD$_n$ can decide $\Pi_n\quad \Sigma_n$ but not $\Pi_{n+1}\quad \Sigma_{n+1}$.

This shows how neatly the SAD machines map into the Kleene arithmetical hierarchy.

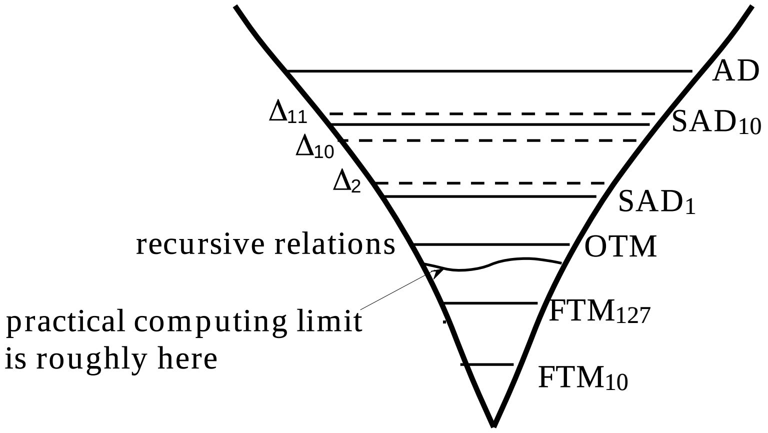


Figure 47. Function space in the light of proposition 3.12.4. Filled lines mark the power limit of computers. Dashed lines mark the boundary of arithmetical relations.

The following conjecture concerns the physical realm, but it has obvious pure implications.

Conjecture 3.12.6. The AD machine can mimic any NTM in a relativistic spacetime.

Two arguments support this conjecture. The first is that *if* SAD$_1$s are the basic building blocks of relativistic computers (as they seem to be), and *if* forming strings is the best method of connecting computers together (as it seems to be), then the AD computer is king because it possesses strings of every order. The other reason is that all my various attempts to construct a machine that cannot obviously be mimicked by the AD have failed. For example, one can show that an AD can mimic a string of ADs; it can also mimic a countably infinite number of ADs working 'in parallel'.

Another physical observation with pure implications derives from this observation: no set-up of TMs in spacetime can perform an uncountably infinite (e.g. $\aleph_1$) number of computational steps. *Proof.* Since spacetime is paracompact (Wald 1984, p. 427) it can be covered by a countably infinite number of (i.e. $\aleph_0$) compact sets. But the amount of matter that could reasonably reside in any one compact set must itself finite, for otherwise there would be at least one point where the density was infinite. Consequently each compact set can house only a finite number of FTMs, and so spacetime as a whole can deliver no more than $\aleph_0$ computational steps.

3.13 A new concept of Computability

It is as when with infra-red
searching a landscape obscured
by the unaided eye one discerns
the sea.

WILLIAM CARLOS WILLIAMS, *The Mirrors*

Computability is shaped, or seems to be shaped, by this question: precisely which (partial) functions should be regarded as computable? Let us try to consider this question *as if for the first time*. (This is an awkward request, reader, I know.) For an answer we naturally turn to the work of Turing, where we find essentially the following.

> (1) a function *f* is computable if and only if *f* can be evaluated by a TM.
>
> (2) the world or spacetime underpinning the TM of (1) is of a particular kind, of which Newtonian spacetime, which has both space enough and time enough to provide suitable accommodation, can be taken as a representative member.

(What Turing actually did was to argue persuasively for (1); (2), on the other hand, was a tacit assumption. The class of functions that (1)+(2) serves to pick out is of course *TCF*.)

Now consider this alternative 'core' condition:

> (i) A function *f* is computable if *f* can be evaluated by a TM operating in our universe.

The reader should contemplate (i) and decide whether it seems right. I think it does. It is certainly true that (i) is more intuitive than (1)+(2) in the sense that (i) can be obtained by combining a logical consequence of (1) — the 'if' part of the 'if and only if' — with a version of (2) that restricts itself to the world that surely matters most: ours. It is important at this point to appreciate that neither (i) nor (1)+(2) can be 'proved' as such; the best that can be said of such statements is that they conform to one's ideas about what Computability is. What I am suggesting then is that *prima facie* (i) conforms to that idea better than (1)+(2).

So let us assume (i) and examine its consequences in the context of our best theory of spacetime on the large scale, GR. The models of this theory that are consistent with current empirical evidence can, for our purposes, be divided into three kinds:

(a) closed universes, (b) open universes, (c) open universes with the M-H property. Thus unlike (1)+(2), (i) does not lead to a single computer; rather there are a number of models in which TMs perform in different ways. Case (a) corresponds to the FTM, case (b) the OTM, and case (c) the NTMs.

If (i) is a sound core condition then Computability should account for all these cases. But how?

The answer lies in an analogy in the concept of Geometry, as I mentioned at the start of this chapter. Geometry, we recall, began as Euclidean geometry, which was supposed to capture perfectly the geometry of the world. Then in the nineteenth century it was realised that one of Euclidean's axioms, the so-called axiom of parallels, could be violated without producing a logical contradiction or even an obvious conflict with observation. This led such authors as Riemann and Lobachevsky to propose the geometrical systems that now bear their names. The concept of Geometry became, and has remained, two-sided. On one side is *physical geometry* which concerns itself with empirical questions regarding the true geometry of the world; and on the other is *pure geometry* which treats the various geometries as self-contained mathematical systems with incorrigible axioms. Not that the two sides are independent of each other. The physical influences the pure because theories of physical geometry (e.g. GR) constantly throw up new systems which the pure geometers can readily axiomatise for their own use. And in the other direction the pure geometers' investigations can help to shed light on the mathematical structure of a particular physical system.

Computability has an analogous origin and evolution. Computability began life as Turing computability, which was supposed to reflect perfectly the most general computer. But now we see that it is conceivable — indeed quite plausible — that there are a number of different TM-based computers that could exist in our world. (The 'no-supertasks axiom' can be violated without producing a logical contradiction or a blatant conflict with our physical theories.) This reveals the two-sided nature of Computability. On one side is *physical computability* which concerns itself with the computational power of TMs in our world and worlds close to our world; and on the other is *pure computability*, which deals with axiomatised versions of TM based computers (FTM, OTM, NTMs) that are thrown up by these physical and mathematical investigations.

Since we are familiar with the case of Geometry — how to make sensible statements, frame questions, provide answers, etc. — this analogy helps us to grasp this new concept of Computability.

Consider the geometric question: do the angles of a triangle sum to 180 degrees? The answer from the pure point of view would be ‘yes’ in Euclidean geometry, ‘no’ in Reimannian Geometry, ‘no’ in Lobachevskian geometry, ‘no’ in Schwarzschild geometry, and so on. The answer from the physical point of view would, as it happens, be ‘no’ because it is known that spacetime is every everywhere slightly curved which prevents such an equation.

So is the halting problem solvable? (Think by analogy.) From the pure point of view the answer is: by the OTM ‘no’, by the SAD_1 ‘yes’, and so a *fortiori* ‘yes’ by all the more powerful NTMs. From the physical point of view this is recognised as a well-defined problem that is closely connected to the as yet unsettled cosmic censorship hypothesis. Another example: is arithmetic decidable? From the pure point of view: ‘no’ by the OTM and all the SAD machines, but ‘yes’ by the AD machine. And so on.

This approach to Computability saves us from becoming embroiled in the seductive yet misguided argument about whether or not the NTMs are somehow ‘less real’ than the OTM. (I regret that Hogarth (1994) was seduced in just this way.) The answer lies in an analogous question: is Lobachevskian geometry somehow less real than Euclidean geometry?

These ideas can helpfully be pictured in terms of possible worlds. First think of each pure geometry as a possible world with that geometry and physical geometry as the science of establishing the respective ‘distances’ (i.e. degrees of resemblance) from a world with our geometry to these worlds of pure geometry. Geometry thus conceived becomes — informally of course — the set of ordered pairs of the form $(G_i, d(G_i))$, where G_i is a pure geometry and $d(G_i)$ is the distance from G_i to our geometry.

The point of this cognitive exercise is that Computability can be represented in an analogous way, i.e., by the set of pairs $(C_i, d(C_i))$, where C_i is a pure computer and $d(C_i)$ is the distance from C_i to the ‘nearest’ computer in our world.

The picture can be brought down to earth. Imagine a long road along which is a series of shops selling computers. The first few shops stock only computers that are available on Earth in 1996, but further down the road the computers on offer become more and more powerful and esoteric. After a drive of a mile we find a *Googol machine*: i.e. a $\text{FTM}_{10^{100}}$. This is apparently available on Earth in the year 3051. Twelve shops on and we find the *Googolplex machine*: i.e. a $\text{FTM}_{10^{10^{100}}}$. This we are told will never be available to Earthlings. A few doors down we spot an

OTM, and then a SAD_1 next-door. Half a mile on and we find a SAD_2, and shortly after a SAD_{19} turns up. An AD is a five minute drive away.

Though obviously tongue-in-cheek, this illustrates the approach one should adopt towards Computability. A problem's solvability is not a binary property (is / is not) as it appeared to be in the light of Turing theory. Rather there is *structure* or *nature* to a problem's solvability, and to grasp that nature requires knowledge of two things: (1) the type of pure computer that can solve the problem / not solve the problem; and (2) how far (down the road) one has to go to find this computer.

Turing's theory can be thought of as an aspect of this 'many-model' theory in the sense that $(OTM, d(OTM))\ \ \{(C_i, d(C_i)) \mid C_i \text{ is a computer}\}$. As such, the two theories do not conflict.[25] But the new theory is an improvement on the old because it has the potential to provide much richer explanations of solvability. By way of illustration, suppose a problem P is analysed using the many-model theory and found to be unsolvable by a Googol machine, solvable by Googolplex machine, and, *a fortiori*, solvable by the OTM and any NTM. From this, and with a minimal amount of physics, we could go on to say that although P is solvable by an OTM, it is almost certainly unsolvable 'for all practical purposes'. Turing's theory, on the other hand, would register only that P is solvable (by the OTM).[26] The point at which P becomes solvable is untouched. As an exercise the reader might compare the two theories' account of the decision problem of arithmetic.

I said there is no conflict between the two theories, but by now some readers may be wondering about Church's thesis.

Church's thesis[27]

A number of different versions of *Church's thesis* (CT) exist in the literature on Turing computability, but the version I will adopt here is this: the OTM can mimic any algorithm that can be executed 'in a finite number of passes'. The claim is non-trivial because it relates the vague notion of an algorithm with the precise notion of

---

[25] Of course the many-model theory does conflict with the common *interpretation* of Turing's theory that the OTM boundary in function space is in some sense absolute. This misguided claim is seen on a par with say the claim that 'Euclidean is in some sense the only genuine geometry'.

[26] This shows incidentally why Turing's boundary is not fundamental *even* in the sense of acting as boundary between the 'solvable for all practical purposes' and the 'unsolvable for all practical purposes'.

[27] Also treated in Hogarth 1994, but the analogy with Geometry, which came later, has made me see things differently. George Boolos kindly gave me some advice on how best to think about Church's thesis within Turing computability.

a OTM. CT could be refuted, but only by exhibiting an algorithmic counter-example. None has yet been found.

Working on the assumption that CT is true, as theorists normally do, has two important benefits. In the first place it gives one a sense of the (enormous) power of the OTM. And in the second, it can simplify proofs by allowing one to argue in terms of the notion of an algorithm rather than its (often more complicated) corresponding OTM.

Within the many-model theory this version of CT remains a well-posed claim and, moreover, nothing I have previously said goes towards refuting it. (I have exhibited no algorithmic counter-example.) We can therefore continue to work on the assumption that CT holds and enjoy the two benefits just mentioned.

But despite this ‘business as usual’ picture, CT has undergone a radical change. For previously it was central tenet of Computability, whereas now it is a conjecture bearing on only one computer (OTM) within Computability. In short: the truth of CT remains as firm as ever, but its significance has been radically downgraded.

The reduced importance of CT is further brought out by the fact that one can make other quite independent conjectures regarding other computers. For example, I conjecture that the SAD_1 can mimic any algorithm that is executed along a single λ-curve.

(Other versions of CT include ‘The OTM is the ideal computing machine’ (Hogarth 1993), ‘The intuitively defined class of computable partial functions coincides with [*TCF*]’ (Cutland 1980, p. 67), ‘All computable functions are [*TCF*]’ (Boolos and Jeffrey 1989, p. 54). But within the new theory these versions are seen to be either false or ill-formulated.)

The final step

The Euclidean system is a geometrical system by anyone’s lights. But actually this system cannot represent the geometry of the world or even part of the world, because as I just mentioned it conflicts with well-established laws of physics. The Euclidean system is therefore neither actual nor even physically possible (in the sense of satisfying the laws of physics). This raises a question: what makes a system a ‘geometry’? Here is Oswald Veblen’s answer: ‘(B)ecause the name seems good, on emotional and traditional grounds, to a sufficient number of competent people’ (quoted by Ryckman (1994, p. 833)). One might try to elucidate this by saying that: ‘a system is a “geometry” if it is “close enough” to the geometry of our world’ or ‘a system is a "geometry" if it bears "sufficient resemblance" to the geometry of our world'. The idea remains vague, but one can begin to see why the Euclidean system deserves the appellation 'geometry' and why say Peano arithmetic does not (nothing very geometrical here).

This has a curious implication for Computability. I have argued that a Compatibility borne of the core idea (i) is akin to Geometry. But if the kinship is truly close — as one eventually feels it is — then (i) is probably not the best core idea at all; that is more likely to be something akin to Veblen's proposal. That is to say, we should ultimately be saying that a system deserves the name 'computer' if the name seems 'good'. Again, one might try to elucidate by saying that: 'a system is a "computer" if it is "close enough" to the computers of our world' or 'a system is a "computer" if it bears "sufficient resemblance" to the computers of our world'. And again although the idea remains vague one can begin to see why say the OTM and the AD should count as computers (even if the OTM or the AD are not physically possible: again think of Euclidean geometry) and why say the system of Peano arithmetic (nothing computer-like here) or the oracle-machine (no *method* of computation as such) should not.

If this is right, then (i) acts as a kind of Wittgensteinian ladder: one climbs it to comprehend the kinship of Computability and Geometry, but once there the ladder can be pushed away.

The reader is now finally in a position to judge whether the 'incredible shrinking machine' and the 'incredibly fast machine' both mentioned in the Introduction to this chapter should be regarded as 'computers'. I believe they both should, and I would add that each functionally equivalent to SAD_1 (*proof*: stretch out along the λ-curve, etc.).

Summary and final remarks

There are those who would argue against the entire project on *a priori* grounds. Finitists of various persuasions (see Moore 1990) would maintain that a NTM is incoherent because it depends on one or more of the following: a supertask, an infinitely long tape, or operating forever. The OTM shares the last two of these features, but Turing's theory can be upheld by finitists because all the suspicious statements referring to the OTM can be recast in terms of acceptable FTMs. (Roughly like this: problem *P* Turing solvable by OTM means there is an *n* such that

FTM_n solves solve P. P unsolvable by OTM means there is no such n.) Since the same cannot be said for the NTM, it must go.

This is no place to argue against finitism *per se*. If the reader has kept finitist principles throughout these investigations then I apologise for having given, in a manner of speaking, a meat cookery book to a vegetarian. Yet I harbour the hope that some such readers may actually come to question their finitist belief *because* of these investigations. There is after all something unusually intuitive about the infinities involved in M-H spacetimes and their associated NTMs. (Though unreachable, the infinity presents itself as part of the world because it can be 'circumvented'.) Moreover the typical finitist objection to supertasks is that supertasks entail some observer actually completing an infinite number of tasks. But if this is the only objection then the supertasks associated with M-H worlds may not offend because HAL never completes her task. All the same, I accept that this hope to change minds may prove naive.

The main technical results are easily summarised. M-H spacetimes are not globally hyperbolic, do possess a non-compact slice, and are prone to infinite blue shifts. A single TM in any M-H world (i.e. a SAD_1 machine) can decide first-order logic, the halting problem, and single-quantifier arithmetic. A SAD_n machine can decide n-tuple but not (n+1)-tuple arithmetic; so SAD machines map neatly into the Kleene arithmetical hierarchy. Finally an AD machine can decide arithmetic, but there are functions that even the AD machine cannot compute.

Computability takes on a new appearance in the light of these results. In Turing's picture it seemed as if there was a fundamental 'boundary' which separates solvable problems from unsolvable problems. In the new picture, however, this boundary is seen to lie within a series of solvable / unsolvable boundaries, each corresponding to the power of a different computer. Arithmetic — to take a most celebrated problem — lies on the unsolvable side of the OTM boundary (Gödel showed us that) but on the solvable side of the AD boundary. Seen from this perceptive Gödel's result might appear to have less significance than was previously supposed.

As I have tried to argue, this all becomes so much clearer when Computability is viewed, so to speak, through the lens of Geometry.

References

Barrow, J. D. and Tipler, F. J. (1986). *The Anthropic Cosmological Principle.* Oxford: Clarendon Press.

Benacerraf, P. (1962). 'Tasks, Super-Tasks, and the Modern Elastics', *Journal of Philosophy*, LIX, 765-784.

Boolos, G.S. and Jeffrey, R.C. (1989). *Computability and Logic.* Cambridge: Cambridge University Press.

Budic, R. and Sachs, R. K. (1976). 'Deterministic Spacetimes', *General Relativity and Gravitation,* 7, 21-29.

Butterfield, J. (1989). 'The Hole Truth', *British Journal for the Philosophy of Science,* 40, 1-28; reprinted in *The Ontology of Science*, International Research Library of Philosophy, ed. J. Worrall; Dartmouth Publishing.

Butterfield, J., Hogarth, M. and Belot, G. (eds) (1996). *Spacetime*. International Research Library of Philosophy. Dartmouth Publishing. Forthcoming.

Clifton, R, and Hogarth, M. (1993). Review of Christopher Ray's book 'Space, Time, and Philosophy', *Philosophical Books,* April, 123-125.

Cutland, N. (1980). *Computability*. Cambridge: Cambridge University Press.

Davies, P. (1983). *God and the New Physics*. London: Dent.

Earman, J. (1986). A *Primer On Determinism*. Dordrechts: D. Reidel.

Earman, J. (1995). *Bangs Crunches Whimpers and Shrieks: Singularities and Acausalities in Relativistic Spacetimes.* Oxford: Oxford University Press.

Earman, J. and Norton, J. (1993). 'Forever is a Day: Supertasks in Pitowsky and Malament-Hogarth Spacetimes', *Philosophy of Science,* 5, 22-42.

Earman, *J* and Norton, J. (1994). 'Infinite Pains: The Trouble with Supertasks', to appear in S. Stich (ed), *Paul Benacerraf*: *The Philosopher and His Critics*. New York: Blackwell.

Geroch, R. (1968) 'What is a Singularity in General Relativity?', *Annals of Physics*, 48, 526-540.

Geroch, R. (1977) 'Prediction in General Relativity', in J. Earman, C. Glymour, and J. Stachel (eds), *Foundations of Spacetime Theories*, *Minnesota Studies in the Philosophy of Science*, Volume VIII. Minneapolis: University of Minnesota Press.

Geroch, R., and Horowitz, G. T. (1979). 'Global Structure of Spacetimes', in S. W. Hawking and W. Israel (eds), *General Relativity, an Einstein Centenary Survey*. Cambridge: Cambridge University Press, 212-293.

Gerver, J. L. (1984). 'A Possible Model for a Singularity without Collisions in the Five Body Problem', *Journal of Differential Equations*, 52, 76-90.

Grünbaum, A. (1968). *Modern Science and Zeno's Paradoxes*. London: George Allen and Unwin.

Hawking, S. W., and Penrose, R. (1970). 'The Singularities of Gravitational Collapse and Cosmology', *Proc. Roy. Soc. Lond*, A 314, 529-48.

Hawking, S. W., and Ellis, G. F. R. (1973). *The Large Scale Structure of Spacetime*. Cambridge: Cambridge University Press.

Hawking, S. W., and Penrose, R. (1995). *The Nature of Space and Time*. Princeton: Princeton University Press.

Hogarth, M. (1992). 'Does General Relativity Allow an Observer to View an Eternity in a Finite Time?', *Foundations of Physics Letters*, 5, 173-181.

Hogarth, M. (1993). 'Predicting the Future in Relativistic Spacetimes', *Studies in History and Philosophy of Modern Physics*, 24, 721-739.

Hogarth, M. (1994). 'Non-Turing Computers and Non-Turing Computability', in D. Hull, M. Forbes, and R. M. Burian (eds), *PSA* 1994, Vol. 1. East Lansing: Philosophy of Science Association, 126-138.

Hogarth, M. (1996). 'A Remark Regarding Singularities and Predictability', *Studies in History and Philosophy of Modern Physics*. Forthcoming.

Horwich, P. (1987). *Asymmetries in Time*. Cambridge, MA: MIT Press.

Littlewood, J. E. (1986). *Littlewood's miscellany*. Ed. by Belá Bollobás. Cambridge: Cambridge University Press.

Lossev, A. and Novikov, I. D. (1992). 'The Jinn of Time machine: Non-trivial Self-Consistent Solutions', *Class. Quantum Grav*, 9, 2309-2321.

Mather, J. N. and McGehee, R. (1975). 'Solutions of the Collinear Four-Body Problem Which Become Unbounded in a Finite Time', in J. Moser (ed.) *Dynamical Systems, Theory and Applications* New York, Springer-Verlag.

Moore, A. (1990). *The Infinite*. London: Routledge.

Morris, S. and Thorne, K. S. (1988). 'Wormholes in spacetime and their use for interstellar travel: A tool for teaching general relativity', *American Journal of Physics*, 56, 5, 395-412.

Müller-Hoissen, F. (1981). 'Determinism in Space-Time', *International Journal of Theoretical Physics*, 20, 6, 443-450.

Penrose, R. (1978). 'Singularities of Spacetime', in N. R. Lebovitz, W. H. Reid, and P. O. Vandervoort (eds), *Theoretical Principles in Astrophysics and Relativity*. Chicago: University of Chicago Press, 217-243.

Penrose, R. (1979). 'Singularities and Time Asymmetry', in J. Earman, C. Glymour, and J. Stachel (eds), *Foundations of Spacetime Theories, Minnesota Studies in the Philosophy of Science*, Volume VIII. Minneapolis: University of Minnesota Press.

Penrose, R. (1989). *The Emperor's New Mind*. Oxford: Oxford University Press.

Penrose, R. (1994). *Shadows of the Mind*. Oxford: Oxford University Press.

Pitowsky, I. (1990). 'The Physical Church Thesis and Physical Computational Complexity', *Iyyun* 39: 81-99.

Pour-El, M. B. and Richards, I. (1981). 'The wave equation with computable initial data such that its unique solution is not computable', *Adv. in Math.*, 39, 215-39.

Popper, K. R. (1982). *The Open Universe*. London: Hutchinson.

Ray, C. (1991). *Time Space and Philosophy*. London: Routledge.

Redhead, M. (1987). *Incompleteness, Nonlocality, and Realism*. Oxford: Clarendon Press.

Rogers, H. (1987). *Theory of Recursive Functions and Effective Computability*. Cambridge, MA: MIT Press.

Ross, S. (1988). *A First Course in Probability*. New York: Macmillan.

Ryckman, T. A. (1994). 'Weyl, Reichenbach and Epistemology of Geometry', *Studies in History and Philosophy of Modern Physics*, 25. 6, 831-870.

Rynasiewicz, R. (1994). 'The Lessons of the Hole Argument', *British Journal for the Philosophy of Science*, 45, 407-436.

Thompson, J. (1954-55). 'Tasks and Super-Tasks', *Analysis*, XV, 1-13.

Tipler, F. J. (1986). 'Penrose diagrams for the Einstein, Eddington-Lemaitre, Eddington-Lemaitre-Bondi, and anti-de Sitter universes', *Journal of Mathematical Physics*, 27, 2, 559-561.

Turing, A. M. (1937). 'On Computable Numbers, with an Application to the Entscheidungsproblem', *Proc. Lond. Math. Soc.*, 42, 230-65; a correction 43, 544-6

Wald, R. M. (1984). *General Relativity*. Chicago: University of Chicago Press.

Wang, H. (1974). *From Mathematics to Philosophy*. London: Routledge.

Wittgenstein, L. (1921). *Tractatus Logico-Philosophicus*. Translated by D. F. Pears and B. F. McGuiness. London: Routledge.