The Amoeba's Secret

Bruno Marchal

August 12, 2023

Translated into English by
Russell Standish and Kim Jones.

Contents

Chapter 1

Introduction

The mind fits itself like a glove
Buddhist proverb

Mr Edgar Morin, president of the selection panel for *Le Monde's 1998 University Research Prize*, urged the laureates, myself included—all of whom were basking in the happy fortune of seeing their doctoral theses published *chez Grasset*—that we not hesitate to create a presentation bearing full witness to the human and personal. Hearing this decided me to relate the entire story.

I should bring attention to the somewhat special character of the result, but also, and perhaps especially, the nature of the path undertaken to get there. I brought up the question of the work's basis for the first time in 1963, at a school in Brussels. I was eight years old: I do not know if I was particularly precocious or merely highly-strung. In effect, the motivation for this work and, for the research surrounding it commencing in infancy, has always been linked to a fear of death.

There are other more fundamental reasons for my describing this unfolding path, an essential part of which lies in my earliest childhood:

1. The work is essentially multidisciplinary. It lies at the intersection of numerous disciplines—theology, psychology, biology, chemistry, physics, mathematics, information technology—and, suddenly, it is difficult to know where to begin. Bearing this in mind, using children's questioning technique is particularly suitable.

2. I asked myself these questions, but quickly started wondering where the questions came from. A good part of the thesis rests on a process of introspection. We will see how the thesis is naturally self-explanatory,

how it explains its own genesis. This is an aspect made clearer if one follows, even if only briefly, the way of children's questioning. I benefitted from this by bringing out a more polished version of the principal argument. There is no attempt at dumbing-down and the reader can skip any passages she finds too technical.

3. The opportunity to do justice and homage to the great authors and books that stake out my quest, amongst them: G. Ames & R. Wyler, James Watson, Linus Pauling, Michel-Yves Bernard, Lewis Carroll, E. Nagel & J.R. Newman, Jean Ladrière, S.C. Kleene, Bernard d'Espagnat.

4. ... and to recount a *Belgian* story at best, or worse a *universal* story, nothing at all of which is very funny. This story explains why I defended my thesis in 1998 in France. I shall recount these events without hatred or a desire for revenge.

I will tell you in several words even now, the principal result. To begin with, the work presents a proof, which is to say a deductive or, to use heavier terminology, hypothetico-deductive argument. This signifies that there is a hypothesis as much as a "thesis", in the more technical sense of what is demonstrated by this hypothesis.

By way of proof, I expect that if the reader is not fully convinced of the result after studying the work, he should either disagree with one of the assumptions or show an error in the argument. Bear in mind that with respect to deductive work, conclusions never officially carry over to reality. Scientific proofs operate inside the frame of a *postulated* theory. Science is thus always modest on the subject of its applicability to, or its approximation to reality.

The hypothesis is that of Mechanism: the idea that we could be digital machines, in a sense that will be rendered more clearly in due course. Broadly speaking, we might be machines in the precise sense that no parts of our bodies are privileged with respect to an eventual functional substitution. This says that we can survive a heart substitution by the transplant of an artificial heart, or of a kidney substitution by an artificial kidney, etc., inasmuch as the substitution is carried out at a sufficiently fine-grained level. Neither can there be any constraints imposed on the level of substitution chosen. It is important to remember that I am not going to defend the hypothesis of Mechanism. I only want to pose this hypothesis at the outset. It constitutes the predefined frame of the work[1].

[1]Note that the idea of taking Mechanism (or Computationalism) as a hypothesis seems

The discovery described here is that in this case, that with this hypothesis of Mechanism, physics becomes reducible to the psychology of *machines*. The "of" should be interpreted in both transitive and intransitive senses: clearly it is a question of a psychology *concerning/about* machines as much as a psychology *inferred* or *postulated* correctly (by definition) by the machines themselves. We will be able, with a little information theory, to define this psychology in the wider sense of the machines' "self-referentially correct" discourse. Such a psychology appears non-normative: we will see that it makes us into beings vastly less well-known to ourselves than we had previously ever imagined. It constitutes a sort of "vaccine" against the numerous forms of reductionism to which human psychology regularly falls prey.

The reduction of physics to psychology happens also at the *epistemological* level: physics effectively becomes a *branch* of psychology—the science of *observable machines*—as it does at the *ontological* level: matter or the appearance of matter emerges from consciousness, from the mind or the mental or even as we will see, of "possible plays/bets/wagers/gambles" made by all digital machines.

It seems that what I have succeeded in demonstrating is that what it is to truly take seriously the hypothesis that we are digital machines is to be forced to recognise a falling-from-grace of the naturalist or materialist idea, quite widespread among philosophers, physicists and the man in the street, that physics is the fundamental science to which all the other natural and human sciences—at least ontologically, and thus in principle—should be reducible. I summarise this theorem by:

$$\text{comp} \rightarrow \text{reversal}$$

where *comp* designates "computationalism", a name often given to "Digital Mechanism" and *reversal* designates the reversal of psychology with physics. What results is not a primitive matter with consciousness emerging from its organisation but the reverse: consciousness is now the more primitive and

to be rather curiously, original. Since Descartes (and even before, notably among the Hindu logicians), there exists a staggering amount of literature surrounding the question of Mechanism and the mind, but it is always a question of arguments in favour of Mechanism or arguments against it. Many also think that Mechanism is by itself a solution to the mind-body problem. This, I hope is a strong suit of the current work: to show that Mechanism does not automatically resolve the mind-body problem. On the contrary, it necessitates a reformulation *of the problem*, taking the form of a necessary justification of any belief in the appearance of a material world, physical or substantial (to anticipate in one phrase the principal result of the work.)

matter, or rather the appearance of material organisation, emerges from all the possible experiences of all the possible consciousnesses. This it does in a sufficiently precise sense that derives physics (science of matter) from psychology (viewed as a very general science of conscious experience, or more positively, of stable discourses by machines themselves: physics, but not geography[2], belonging necessarily to this self-referential discourse, which I will demonstrate.)

At this stage, anyone who for any good reason were to be persuaded of the veracity of contemporary materialism[3], can always surmise that the present work constitutes a rejection of Mechanism. This will nevertheless pose a problem since Mechanism is, implicitly or explicitly, the philosophy adopted by the majority of materialists.

Concerning my own position on this I remain silent. My philosophical opinions rest—and will remain—private. In the more technical part of the thesis, I nevertheless show that one can already extract enough qualitative and quantitative givens from physics once this is shown to be derivable from machine psychology. We can then confront the results with the usual empirical and modern physical theories—notably quantum mechanics—to start to see an empirical confirmation of this psychology and thus confirmation of the reversal.

By illuminating problems in the interpretation of (quantum) physical facts, this thesis leads to a de facto judgement: that the reversal and its reasoning logic, Mechanism, are plausible.

A final observation concerns rationalism and interdisciplinarity.

This work pleads as "rationalist". Like Karl Popper, I appreciate the contrast of rationalism and elitism. Rationalism is a form of hope concerning the reasoning powers of others. It is the hope that the other will have the courtesy to listen to you and accept your results or to indicate to you your errors, or to say to you at the very least that the subject is of no interest to him. Popper writes:

> Faith in reason is not only a faith in our own reason but also—and even more—in that of others.

[2]Physics becomes the study of what is *a priori* observable by every observer. The moon's existence is not (in all truthfulness) a physical law. By now one might fear that the physical laws lead only to trivial truths, but we will see that the constraints of Mechanism de-trivialise this introspective physics.

[3]Throughout this work, "materialism" will be taken in the weak sense of the philosophical doctrine that postulates the existence of a substantial universe (of things obeying laws independent of us).

> Thus a rationalist, even if he believes himself to be intellectually superior to others, will reject all claims to authority since he is aware that if his intelligence is superior to that of others (which is hard for him to judge), it is only insofar as he is capable of learning from his own and other peoples' mistakes, and that one can learn in this sense only if one takes others and their arguments seriously. Rationalism is therefore bound up with the idea that the other fellow has the right to be heard, and to defend his arguments. (Karl R. Popper[49])

I feel in particular, that reason is a universal value and, universally profitable. Nothing else like science exists that is clearly separate to all other human endeavours. I happen to believe that there are those possessed of a scientific attitude, which is no more than a disposition toward modesty and honesty with themselves and with others. This attitude does not depend on any particular domain. I have a ready-made slogan:

> Some gardeners are more scientific than some astronomers

And, I might have said *astrologers* in place of gardeners[4].

Today a kind of artificial chasm is maintained between the human sciences and the exact sciences. To combat the so-called *elitist* usage of mathematics, a minister of politics[5] was struck by the notion of suppressing numerous hours devoted to maths in diverse sections of secondary teaching.

In the same way, less mathematics is being taught in social science. This can only end up discouraging teachers from teaching by demonstration—ie explanation—of formulas in mathematics courses. Clearly, the opposite is needed, to give up teaching altogether of other mathematics in social sciences.

The prohibition of the generalised use of deductive or interrogative reason and of mathematics, has contributed not only to rendering the social sciences less exact and the exact sciences less human; it has also especially contributed to rendering the social sciences less human and the exact sciences less exact—as should be clear from the present work.

Note that I do not claim that reason is everything, or that it is a kind of universal panacea. I only say that reason, properly considered, should stand as the first rung of courtesy, permitting evaluation and progress in

[4]One can consult the fine book by Suzanne Blackmore *In Search of the Light*[7] for an example of a scientific contribution to parapsychology, albeit somewhat negative.

[5]This idea was defended by Claude Allègre, explored in his book The Defeat of Plato [2] and applied when he ascended to the post of Education minister.)

the research of knowledge and allowing courageous "backflips" from time to time, as in revising one's beliefs or abandoning a prejudice.

Reason is not sufficient for progress in knowledge. There must also be inspiration, attention, imagination, bravery etc. While reason is not of itself sufficient, it is necessary though to communicate results to others.

Again, concerning interdisciplinarity, I often like to cite Descartes. He wrote:

> One must therefore be convinced that all the sciences are so linked together that it is easier to learn them all at the same time, than to isolate each from the other.

I hope that the present contribution will illustrate to what extent Descartes was inspired on this point. As with the collaborations on quantum mechanics of Einstein, Podolsky and Rosen on the one hand, and of Bell on the other, this work should also illustrate the artificial character of the border dividing science and philosophy, or even between science and theology. We will return to this point.

Any trace of eventual barriers existing between the sciences and philosophy rests ultimately on philosophical postulates, avowed or not. The advantage in my briefly recounting to you the pathways taken by my burgeoning youthful thoughts, lies in the fact that children are naturally interdisciplinary: they have not yet submitted to that form of brainwashing known as "academic specialisation". Children will always pose questions without fear of where they might be putting their feet.

Chapter 2

The Amoeba's Secret (1961 → 1971)

what am I doing here in these miasmas
tiny little Lilliputian
seized by terror, sometimes by asthma
before these tonnes of thingumajigs
Gaston Compère, "Geometrie de l'absence"

What follows evidently constitutes a partial view of the past. I am not telling my life-story but calling on those sparse events that illustrate the threading of the ideas and questions that together form the origin of the discovery.

Certain paediatricians claim that the first *metaphysical crisis*, or the first anxious moments concerning death, occur in children around age 4. Perhaps. I remember well the terror that invaded my mind day and night, and I demanded all manner of assurances from my parents that I would wake up the following morning.

With the well-intentioned care of silencing any woes in children, parents are apt to tell stories. As I was born in Germany, a nanny would regularly read to me in German many folk-tales, mainly those of the Brothers Grimm, although I am not so certain that these offered me any appeasement concerning my worries.

It must have been at around the age of 5 or 6 when as I remember, doubtless by a sort of absent-mindedness or simply out of fatigue, since I assailed him with questions remorselessly, that my father informed me that Saint Nicholas did not exist.

"And Father Christmas?"

"Him neither" my father replied, sadly noting my incredulous and thunderstruck countenance. Thus collapsed my first theory—or ontology, mythology, theology, belief, dream... call it what you will; at this age, any precision on my part had been premature.

"And the fairies?" I pleaded.

"Them neither"

"But, come on—the angels and all that...?" Here I could see that I had come once again to pose my father an embarrassing question. After a long drawn-out sigh, he explained to me that really he believed in none of it—neither in the angels or in God, but that my cousins and uncles and aunts believed all of it. This astonished me all the more in that fairies were for me no more than female angels equipped with magic wands. I liked fairies and angels because they could fly (had I gone on to develop this tendency I might have become an aviator), but in the main because fairies and angels were immortals.

By now I had realised that adults could have differing beliefs. I found this profoundly shocking. If my cousins could believe in angels, was it not my *right* also to believe in them, as well as in fairies?

My father explained to me that it was surely my right after a certain fashion, to believe in whatever I wanted to believe, but that it was by no means evident that to do so would be in my best interest.

To believe in false propositions is to invite deception and disappointment. I found entirely pathetic the notion of believing in the false, and the whole thing gave me the shudders. From this moment on, I would try to adhere to the rule: avoid at any price belief in falsity.

The truth, evidently, can give rise to fear. In particular the idea that I was a mere mortal seemed to me to be at the very limits of the acceptable. But the idea of believing in falsity out of fear of the truth worried me all the more. I therefore made a promise to myself to always search for the true, fearsome as this may well turn out to be. To know would seem even better.

To know is better: agreed. But is this even possible? Surely it cannot be easy.

To start with, I observed that during nightly dreams I was able to believe in just about any falsity. In addition, I suffered sleep problems, like many kids, something confirmed by electroencephalography. My dreams were abnormally realistic. This hyperrealism was fine in the case of lovely and pleasant dreams, but it became truly worrisome in the case of strange dreams and nightmares. Doubts arising from dreams, even concerning the possibility of knowing truth, will play a role in the story that occupies us here. There is nothing original in any of this; the metaphysical role of dreams

appeared with the Hindu idealists, Plato, Descartes, Berkeley, as I would learn later on.

There followed the problem of a divergence of opinion between my father and my uncle. Before, everything was simple: a proposal was true if and only if my father asserted it[1].

Since he had evinced several seconds of doubt over the existence of angels and told me that my uncle himself believed in them, I truly wondered just whom I should believe about this.

I asked my uncle why he believed in angels. He replied—inasmuch as I can recall with any precision—that his belief was based in the fact that his parents believed in them, also his grandparents, etc. I found his reply frankly troubling. In effect, if his ancestor had been mistaken, this mistake would be propagated from generation to generation. I came to admire my father's placing in doubt his own parents' beliefs, and I decided never to believe in a proposition under the mere pretext that it had been pronounced by a trusted person or family-member. I had hit upon what one now calls the principle of free enquiry, a founding principle of l'Université Libre de Bruxelles, the university where my father concluded his legal studies after studying with the Jesuits. I have no doubt that he may have influenced me.

I asked my father why he did not believe in angels and fairies (I could not have cared less about St Nicholas and Father Christmas because to my mind they were not even immortals). He responded by saying that having thoroughly searched everywhere, no one had encountered them anywhere. There followed a deluge of revelations: we live on a ball suspended in space, we have already orbited it etc. It seemed like no place existed for fairies and angels.

In order to not run the risk of believing falsity, my interest in *imaginary* beings slid over to a pronounced interest in animals, for whose existence nobody had even the slightest doubt. Returning to Belgium from Germany, my parents bought a small hobby-farm in the country to which we would go during vacations and on the weekend. I passed a lot of time observing swallows, butterflies, ants etc. When observing an animal, for example, a butterfly, I identified body and soul with this butterfly. If it flew, it was I who flew, if it gathered pollen, it was I who gathered pollen and it was I who became intoxicated by the multiple nectars of the flowers of the fields.

One day, I pointed to a white butterfly and exclaimed to my sister and brother "Look! This butterfly, I recognise it, it's me; I have been this

[1] I express myself here in adult language; at the time I would have been hard-pressed to formulate such a proposal in this way.

butterfly for several weeks now." And they, with a delicacy well known amongst siblings, broke the news to me that this was not possible "because butterflies only live for one day."

This came as quite a shock. It reminded me that if swallows and butterflies flew, like angels, they were none the less mortal for it, like me. They did seem though, to live a much shorter life than I, which I found disturbing.

At this time, whenever I identified with an animal, the identification took place in real time: I did not yet imagine from the butterfly's point of view that one day could appear very long. Thus, if a butterfly truly lived but one day, due to my identification with it, I also lived but one day and no longer. And this was no laughing matter. I became maniacally obsessed with the maximum life spans of animals. Every time I heard of a new animal I would ask about its longevity. I was rather disappointed to discover that, on the whole, large animals lived longer than the small ones with which I had been identifying almost exclusively since I myself was of small stature at this time.

It was then that I made an authentic and revolutionary discovery. I had a canine companion in whom I confided my metaphysical concerns, my partner in the quest for truth. One fine day, I tried to show him a tiny red spider (in fact a tiny garden *acarina*), without managing to attract his attention. I concluded that the *acarina* was too tiny for my dog to see and suddenly, I found myself identifying with my dog. It thus came into my mind that the fairies and angels were perhaps just that little bit too miniscule for us to be able to perceive them.

As quickly as I could, I presented my theory to my father. I was particularly serene, not only in view of a proof of the existence of fairies, but also of the proof that my father could not be sure of their non-existence. I played devil's advocate not because I wanted to contradict my father at any price, but rather to show that my cousins and uncle were perhaps not entirely in the wrong.

"Even if you searched everywhere on Earth for angels and didn't find any, that proves nothing" I said. "Perhaps angels and fairies are simply too small for us to see them?" I explained to him the experience with my dog. My father, who had an answer to hand for everything, enlightened me that the search had included the direction of the tiny as well. He spoke to me of the microscope and—and in fact it was this that surprised me the most—he explained that the effective discovery of a multitude of tiny animals *invisible to the naked eye* had thereby been made. He then took a piece of paper and drew a sketch of an amoeba. I fell headlong in love with this tiny and adorable creature, multiform and so easy to draw.

And so to the fundamental question of the moment: how long might an amoeba expect to live?

Considering my belief that the smaller an animal was, the less time it could live, I hardly had any illusions. It must be that my tiny amoeba could not possibly live very long at all.

On this question of the lifetime of an amoeba, my father, with infinite wisdom, contented himself to explain that having eaten their fill of even tinier (!) creatures during the day, rather than merely dying like any ordinary beast such as a butterfly, it divided itself instead into two. Instead of dying and disappearing, an amoeba would divide itself and become two amoebas. This was practically the reverse of death itself.

"So they're immortal, then?"

This time, my father made no response.

I requested, of my elder brother and sister notably, that they bring back from school as many documents as they could find on amoebas, which they very kindly did. I thus started to write (more exactly, to scribble in just about every sense) a book: *The Invisible World*. My idea was that if invisible worlds existed—and the existence of the amoeba proved the existence of such worlds—one could no longer enjoy any form of certainty over whatever the case might be. In the final analysis, my uncle may well have been right on the subject of angels. The amoeba was surely a tangible piece of evidence that at least certain animals could be immortal. I killed time by annoying my parents with the demand for a microscope of my own. When the microscope inevitably arrived, I looked for amoebas. I discovered euglenas and especially paramecia, and when they divided themselves into two, I divided into two also. The question now was to know whether the paramecium had survived its division.

What exactly was going on?

I arrived at my first public seminar on amoebas. Even though I may be driven by self-imposed questions, I have always had an immense enthusiasm for giving oral exposés, even delivering classes and seminars on subjects at a considerable distance from those that preoccupy me directly. Thus, I had already given several verbal presentations, notably on minerals, but, in 1963, at the age of 8, people were urging me to deliver a seminar on *microbes*.

Entitled *Amoeba, Euglena and Paramecium*, I have managed to rediscover my succinct resumé in an old notebook:

> My friends let me tell you, in this room, we are not 24 in number,
> but several million [2].

[2] I already liked paradoxical propositions; true but slightly astonishing statements, so to

> Does the elephant see the tiny red spider? Could living beings exist who are so tiny that to us they would be invisible? Could there be an invisible world and a tunnel through which to explore it? As incredible as this may sound: yes. The microscope is the tunnel and the microbes are the discovery. Amoeba, euglena, paramecia, vorticellae, stentor, bacteria, ovum and spermatozoa, the protozoa among us! Nutrition, digestion, excretion, diverse sensibilities (the euglena's eye), and ... *reproduction*.
>
> Question: How long can an amoeba live? One day or forever? If it lives two days it lives every day... forever. (Robert Catteau public secondary school, in the presence of Prof. Verschaeve)

I would become more and more obsessed by this question of the immortality of the amoeba. For the following two full years, I would pass half my free time on walks gathering every possible kind of water (sewerage, liquid manure, pond water, estuarine, puddles of every kind) and the other half observing these waters under the microscope. As usual, I always identified completely with the micro-organisms I was observing and attempted somehow to *sense* whatever was going on at the moment of their division. I scaffolded an unimaginable number of theories illustrating the immortal character of unicellular creatures without arriving at a stage of conviction over any of them. The consistent effort to go from fairies to amoebas had been kick-started by my fear of believing in non-existent things and I did not at any price want to believe that amoebas were immortal if they in fact were not.

Yet, certitude had come through: IF an amoeba lives two days THEN it lives every day[3].

It remained to demonstrate that for an amoeba to be immortal, it only needed to survive one division.

An example of a theory heading in this direction was what I called "The Principle of the Inspector":

NO CORPSE means NO MURDER

speak. "We" to my mind, evidently designated the students of the class with the teacher and the microbes in the class. The "million" would have had to be a much higher number in reality, if one had wanted to be more exact.

[3]In fact the common amoeba divides on average every 50 hours approximately, but for the sake of simplicity I will continue to speak as though its divisions occur every 24 hours.

According to the inspector, when the amoeba divides, it leaves behind no corpse, so no "body" dies in the act of division, therefore an amoeba survives its division. But this reasoning is invalid. When a hydra eats an amoeba, it gets digested and neither is any corpse left behind. The difficulty lies in believing that it survives the process of digestion. The "Principle of the Inspector" collapsed.

My basic theory or argument in favour of the immortality of the amoeba or paramecium had been directly linked to the experiment of swapping places with a concrete paramecium and keenly observing it through a microscope. Evidently I had come up against a problem of scale. It seems that I go from one to two but how is this possible? Which is the original paramecium between the two new ones? Both or only one of them? Which one?

In particular, if I become one of the two paramecia, how could I convince the other, given that it could just as easily make the claim I am making: of being me?

Completely gobsmacked by this, given over to a sort of semi-ecstatic vertigo I realised something just as extraordinary and incommunicable.

My feeling had been that the amoeba survived its division (and thus every division, meaning that it was immortal) but as it had become two, each of the two resulting amoebas was now unable to convince the other that it had survived, where "it" referred to the original amoeba. From whence arose the incommunicability.

If an amoeba could not bring any of its copies to accept its survival or its immortality, how much more difficult might it be to convince a human being?

How difficult might it be for me to convince another human of the immortality of an amoeba even if immortality were an accepted notion? The more I reflected on this, the more it seemed to me that this immortality, if such it was, must be condemned to remain forever secret. I had explained to my satisfaction the prudent silence of my father.

A spectacular confirmation would arrive when I was given *The Marvels of Life*, the very fine book by Ames & Wyler with a preface by Jean Rostand and superb illustrations by Charles Harper—it was also the first book I ever took to bed and slept with!

This book contained an entire chapter consecrated to the amoeba. Distraught by the toll of new information it contained, I initially believed that it would not get down to the question of the immortality of protozoa, but one day, I fell upon the photo of a paramecium for which the legend was "Is the paramecium immortal?" I quickly felt relieved because I could see that one *could at least pose this question*. Soon thereafter I felt astonished:

"here, finally is a book that addresses an enormous number of questions and is content to pose *the* question". This astonishment was given legs by the confirmation that the immortality of the paramecium, if immortality exists, could be no more than an unavoidable query: a wager on uncommunicable success.

Ames and Wyler were just as prudent as my dad. I wondered if I was going to succeed at being just as prudent as them. What rotten luck all the same: I discover a fundamental truth and it seems forbidden to communicate it. I would have to wait until 1971 to get out of this impasse and to weigh up the communicable as per incommunicable parts of the amoeba's secret.

It is noteworthy that up until this time, I had never asked what an amoeba was made of, or indeed what I myself was made of. It seemed to me that the question did not truly depend on whatever things were made of. I did not imagine myself as made of something(s). The problem of immortality seemed to me to be more a question of biology, or of psychology, or of theology—not a question of physics. The matter did not rest there either in that I demanded to know how an amoeba managed to divide itself into two. Moreover, the argument in favour of the amoeba's immortality on one hand and especially the incommunicability of this immortality by the amoeba on the other hand, depended crucially on the fact that after the division, the two resulting amoebas were rigorously identical, since only in this case did the two amoebas seem to contradict one another in claiming to have survived, each, both!

The reading of Ames & Wyler, accompanied by books of Jean Rostand passed to me by my father as well as some excellent manuals of Jean-Pierre Vanden Eeckhoudt—teacher at the Robert Catteau Public School—would drive me from astonishment to astonishment. There, I learnt the magic words that described the principal phases of cellular division: prophase, metaphase, anaphase, telophase; as well as their significance in chromosomal terms. I learnt especially that I am myself constituted from a colony of social amoebas! Our pluricellular organic quality posed me problems: how could I as a society of amoebas still identify with an individual amoeba? Unless an amoeba itself were in turn a colony of sub-microbes, and so on and so forth? I was thus led naturally to an interest in chemistry and to atomic physics.

On the subject of matter, I would pose myself a question that would prevent me from truly taking seriously the idea of the atom, to say nothing of the very notion of matter itself. Initially, I imagined atoms to be ultra-smooth and ultra-hard spheres; next I learnt that atoms were in fact constituted of electrons spinning around a nucleus of protons and neutrons that I imagined were in their turn, like ultra-smooth and ultra-hard spheres.

It seemed that you could always divide matter and that research to find the ultimate particle was all in vain. At the same time however, if one were to find an ultimate particle, it seemed to me, what could it possibly be other than a smooth and ultra-hard sphere once again and forever, and what could such a sphere be made of?

The very concept of matter seemed to me to be devoid of any explanatory capacity. To me, the notion of matter seemed to toss out more questions than it did answers and seemed even to threaten the unity of the amoeba.

It was in the volume by Joël de Rosnay that I learnt of the existence of DNA[4], the gigantic molecule of deoxyribonucleic acid which is a long chain in the form of a double helix "like the inside of a certain castle of the Loire", comprising the repetition of molecules taken from the group Adenine, Thymine, Cytosine, Guanine resulting in a very long "word" in the genre of AATGGCTATGGACCTCAG... and it was in this book that I would learn how this word, seen as a suite of triplets AAT GGC TAT GGA CCT CAG.... is translated into RNA, another nucleic acid, itself translated into a "word": proteins, comprising tiny molecules, amino acids, chosen from the alphabet of 20 "amino acids". I would learn how these proteins and their enzymes coped with the rest: from the synthesis of tiny molecules (amino acids, nucleotides, sugars), even right up to the constitution of the cell.

All the same, this gave oxygen to more questions. How did we know all this? What, indeed *is* a molecule?

In fact my "Joël de Rosnay" and the review *Science & Life*, was a springboard for the book destined to become my basic bible for the following years (1968 and thereafter): the French edition, edited by François Gros and prefaced by François Jacob of James D. Watson's *The Molecular Biology of the Gene*. In this book, I would gain a glimpse of the incredible molecular dance that goes on, not only with the amoeba, but also with an even tinier creature: the bacterium *Escherichia coli*.

My "Watson" was so *biblical* that in my vocabulary, the very word "Watson" had become synonymous with THE Bible. In retrospect, my "Ames & Wyler" had been my first *Watson*, but at the age at which I read it, I believe that I did not pose questions in the genre of knowing who might have written a book.

Even though *The Molecular Biology of the Gene* enjoyed pride of place as *the* Watson, another Watson rapidly outflanked it: *General Chemistry* by

[4]Of course Ames & Wyler also speak of this, but I had no real understanding of what was in question. More and more, my Ames & Wyler opened automatically to the little chapter consecrated to the amoeba!

Linus Pauling, and this augmented a bitter conflict in my mind, destined also to endure through the years to come.

On the one hand, my Watson gave me the impression of a wonderful molecular dance perfectly encoded by RNA and decoded by the cell. This allowed me to see the essential machinery, wherein each molecule's identity made no contribution to the identity of the whole organism. All of this being accounted for, it now made sense to say that the amoeba, repeating in a fashion much more complicated, though in principle similar to the molecular dance of bacteria, reproduced itself *mechanically*. I was able to observe, through molecular genetics, an *implementation*[5] of a solution to the problem of knowing how an amoeba could manufacture another amoeba identical to itself.

Concerning the lifetime of an amoeba, this proceeded in the direction of an amoeba's survival of its duplication (a high-fidelity reproduction). Thus, given the "basic theory" investigated in 1963 [6], an amoeba is immortal.

This is truly independent of the fact that the amoeba cannot communicate this, and just as independent of the fact that "I" cannot communicate it either.

My Linus Pauling was a kind of concrete proof that I was somehow forbidden to relate the truth about the amoeba's immortality. Watson said, "Cells obey the laws of chemistry". To be sure of the truly "machine-like"—discretely causal if you will (I had yet no concept of the *digital* or the *numerical*)—character of self-reproduction, it was imperative that I assure myself of the discrete and machine-like profile of the activity of molecules themselves.

Now, if Linus Pauling abounded along the lines of the discrete aspect, with quantification by nature's chemical properties, including atoms and distinct energy levels—all seemed to rest on mathematics, perhaps even on the real numbers; on the continuous, the differential equations. How about that! What was I dealing with?

The conflict between Linus Pauling and Watson was for me a truly palpable war of ideas, which took on major proportions at the onset of the student vacation. As always, and to my lasting joy, we headed off to the countryside, and the question loomed: do I take Watson or do I take Pauling? I knew full well that if I took both I would pass the entire vacation vacillating between them. As it was during the school year, I would dissipate

[5] a representation in terms of data; also known as an "implantation"

[6] An amoeba lives one day or every day. Otherwise put: if an amoeba lives for two days it lives forever. Or again: if an amoeba survives one of its cellular divisions, it survives every one of its cellular divisions.

my time by continually flicking pages in the one and in the other without deciding on anything and remaining stuck in an abyss of perplexities.

This progressed to the point of reading other authors' works, the "non-Watson" literary genre, in a range comprising the diaries of *Tintin and Spirou* on which I was fixated, whodunnits, Freud, Young, Ionesco, Borgès. There was even a certain *Alice in Wonderland* with which I became bored to death and though I gave up on it halfway, I would nevertheless return to it later on.

During this whole time, I had kept my taste for oral exposés. In Biology class at high school, I constructed an exposé on "the lactose operon" (results due to Jacob and Monod in bacterial genetics). Though I had not finished my talk, the teacher let me continue the following hour and again the following hour etc. In the end, he allowed me to have my say during several weeks. One day, the teacher summarised my exposé. This had evolved into a small introduction to molecular biology and during this, he committed a negligible error. Always careful vis-à-vis the truth, I discreetly informed one of my classmates but one of them (the traitor) pushed me up against the speaker's platform saying "Sir! Marchal wants to tell you something." I was thoroughly annoyed by this and with infinite *politesse* let the teacher know about the error in his summary. He remained silent for a while, deciding ultimately to put all the students straight on his mistake. He was never the type to hold out on me and we developed a relationship based on mutual respect. In fact I have an enormous respect for those who can recognise and correct their errors. Recall how much I admired my own father for changing his opinion. With Camus, I opine that perhaps the only eternally persisting thing is stubbornness.

I fretted also over the choice of university studies even though this still seemed a long way off. I was nevertheless extremely impatient to get to university if only to be in a position to pose all those questions that both excited and oppressed me. Would I do biology or chemistry? I asked myself this question practically every day.

By then, I would have the luck to be able to frequent the Molecular Biology Laboratory of ULB at Rhodes-St-Gènese, thanks to the kindness of Jean Rommelaere, whose mother was a friend of my own mother. There, I would meet Jean Brachet who was director of radiobiological services and I especially got the opportunity to meet and chat with René Thomas who directed the bacterial and viral genetics service—Escherichia coli and the *lambda* phages. What an opportune encounter! René Thomas was the biologist who discovered the formal logic in Lewis Carroll's book *The Game of Logic*. This was a wonderful book the French edition of which contained

magnificent illustrations by Max Ernst, including a drawing of a marvellous planarian worm. Most interesting in René Thomas' work was his showing how logic circuits could be simulated in bacteria by means of genetic monitoring of the genome of the bacteria; thereby corroborating my intuitive reading of Jacob and Monod's article to whit life was a matter of encoded dialogues. We promised each other to meet up again; I still had 2 or 3 years of high school to go. This encounter unleashed my impatience to get to university.

I continued however, to pursue my interest in chemistry and in the question of the amoeba's constitution. From Linus Pauling I would move on to my next *Watson*. A real tiny masterpiece came my way: none other than Michel-Yves Bernard's book *Introduction to Quantum Mechanics and Statistical Physics*. It was a rare introduction to quantum mechanics written for secondary school students.

Finally, I fell back on the question "what is matter?" Organisms are societies of cells, cells are societies of molecules, molecules appear to be societies of elementary particles, but the relation between the particles seems to necessitate a science of the continuous. But—what IS the continuous? What, in addition, would the relevant advanced mathematics bring to the issue?

To summarise, biology and molecular genetics presented strong clues that we are machines. Our biological identity seemed to me to be defined by the encoded information and essentially independent of the material involved, this being continually replaced. In this case, the amoeba faithfully reproduces itself and must be immortal since its identity resides in its form and activity—and not in its substance. Hidden behind this, chemistry and mathematics throw a shadow of doubt on this mechanist conception. Even in Newton's "mechanics", objects—often identified as "material points"—seem to act at a distance by means of scalar fields in space and described by a mathematics causing the intervention of the mysterious continuum. With quantum mechanics, this aspect of things seems pushed to extremes: even an isolated particle or an atom are described by functions that only cancel out at infinity. It wasn't at all evident to me how, under such conditions, the amoeba could make identical self-reproductions, or even how it might otherwise self-replicate. With quantum mechanics and the continuous, it seemed that a filament might always subsist between the two apparent amoebas and that in reality, there existed but one amoeba making a good impression of seeming self-division.

In 1971, on the eve of a scholastic voyage to London, the spiritual conflict between biology and chemistry hit its apogee.

Above and beyond the immortality of the amoeba on one hand, it seemed that the explanatory power of molecular genetics resided entirely in Digitalism. This permits the use of encodings and provides an explanation in quasi-psychological terms: of memory and its transformation and interpretation. With hindsight, the article of Jacob and Monod on the Lactose operon—rehashed by Taylor[60]—represents my first discovery of the formal explanation of "IF... THEN... ELSE" of logicians and computer scientists.

On the other hand, this quasi-psychological explanation of the functioning of the cell seemed vastly incomplete without a clarification of the nature of matter.

It sufficed not to say simply that there are things obeying laws; one must also explain what these things are, where they come from, why they obey laws and where the laws themselves come from.

"Cells obey the laws of chemistry," said Watson. We shall see. Might it even be that chemistry obeys the laws of cells, as though chemistry were the product of an amoeba's dreaming....

Chapter 3

Gödel's Diagonal (1971 → 1973)

If DA gives AA; and DB, BB; and DC, CC; what gives DD?

In 1970 I enrolled in "Poetry". This is the penultimate year of secondary school. The final year is called "Rhetoric". My impatience to get to university was such that I put myself down for the central jury examinations with the idea of leaping over my last two years of high school. Ultimately, I did not pursue this enterprise, a largely paradoxical affair. Not only was I effectively in a constant state of hesitation between biology and chemistry, but my doubts had also enlarged to the point where I could now see myself opting for philosophical studies.

As an independent student, I would attend different classes at university by cutting a few hours of classes at school. In particular I attended the exciting chemistry classes of Lucia de Brouckère from whom I gained not only the brief opportunity to go over my hesitancy, but also with whom there was time to chat in a spirit of free enquiry. Lucia de Brouckère was a towering figure of secularity and freedom of thought in Brussels. I continued to go to the Molecular Biology Laboratory at Rhodes-St-Genèse, although now I only bothered with its library.

Irritated by my own hesitancy as I said earlier, I ended up reading all sorts of books, most found at random during a walk through a bookshop. It was through a reading of Gilles Deleuze's book *Logic of Sense* that my mind was finally opened to Lewis Carroll and especially his book *Sylvie and Bruno* that I read several times in quick succession. I again took up *Alice in Wonderland* and also *The Game of Logic*. I still—even today—manage to maintain the claim that English humour is built on the taking

seriously of classical logic; this never works—which explains the inherent laughter-value. As a result, I started to take an active interest in logic and in paradoxes of ensemble theory. Besides, I knew that Chemistry brought advanced mathematics into the picture, and I had to admit that I took great pleasure in Maths classes at school.

During my scholar's voyage to London the year before, and in Amsterdam, I recall buying only (Anglo Saxon) books on genetics and chemistry including William Hayes' fine book on bacterial genetics and their viruses[31], as well as the Taylor[60] that contained the article by Jacob and Monod. I penned an enthusiastic letter to Bill Hayes who responded with fervent sympathy. That year, now in "Poetry", I was going to London[1]. The conflict between chemistry and biology was at its maximum and in Foyles' Bookshop, I fled this internal quarrel by giving myself over almost entirely to the range of Lewis Carrolls as much as to the mathematics and logic sections of the store.

It was there that I came across the little red book *Gödel's Proof* by Nagel and Newman. I had no idea who Gödel was and with what his proof was concerned, but in browsing this book I took from it that the work presented a proof on the subject of the existence or the in-existence of a proof. I was intrigued. I next understood that this state of affairs had been achieved by means of an *encoding*. The resemblance to biological encodings literally leapt off the page to my eyes.

Without necessarily believing in it too much, I progressively realised that this piece of work proposed a general means to facilitate the construction of formal expressions[2] capable of referring to themselves. I was surprised to discover that these expressions were perfectly well defined by the signs and symbols that represented them, in much the same way as the amoeba seemed to be defined by the molecules and atoms of which it was constituted.

I had a good idea of how the amoeba or *Escherichia coli* self-divided, which is to say, I had a quasi-visual model of reproduction at the molecular scale. I thus had to hand a kind of proof that the amoeba could clone itself; but as it turned out, this model (as I have already mentioned) rested on the manner in which the molecules interacted. Because of this, I could not be sure that the amoeba had even replicated—either by itself or with an exact

[1] I so appreciated London and Oxford for their scientific booksellers—and Lewis Carroll—that from that date on, I would go to England every year, notably to Oxford for what I called my "Carrollian Pilgrimage".

[2] collections of signs lending themselves to interpretation at the core of a formal theory, like logical statements, capable of interpretation by a machine, like a software program. At this time, none of this was anything like a clear notion for me.

clone resulting. The amoeba's genome or its genetic encoding had been seemingly replicated identically, but this capacity for identical replication rested to all extents and appearances on the laws of chemistry, which in turn seemed founded on the continuous.

I wondered whether it truly was the amoeba—this tiny and discrete unit with which I identified since my earliest childhood—that divided itself, or, indeed whether it was the very universe itself—which I conceived of as a gigantic and unknowable continuum—that divided the amoeba.

With Nagel & Newman in hand, I started to understand that it was possible to envisage self-reproducing entities having *a priori* no links with chemistry or with the continuous, or even apparently with the universe of physicists or chemists. I discovered a new sort of abstract amoeba that may well be infinitely easier to interrogate than the tiny, concrete critter inhabiting the pond-water of the neighbourhood.

Practically speaking, "Gödel's Proof" seemed to turn the key in the hesitancy lock-up between Chemistry and Biology. This was a veritable triumph for Biology, all the more so in that its transformation into an abstract biology of formal beings—concerning whose nature I yet lacked complete clarity—might be in order.

(The dilemma of knowing whether or not, in this case, I could still identify with the amoeba had yet to fully surface. At this stage however, I was so happy to have discovered a totally new kind of amoeba that I relegated this question to the future.)

There were *other* things about Nagel & Newman! Not so much in Gödel's proof—where apparently self-reproductive or self-referential entities appeared—but in the result, in Gödel's *theorem*, specifically in his second incompleteness theorem, published in 1931. In effect, and in rather crude terms, it seemed that there actually exist widgets, say, with the capacity to communicate true propositions (and me—in love with the true! [3]), capable, in addition, of communicating apparently true propositions concerning themselves, but (it would seem) directly because of this, incapable of communicating or of demonstrating[4] certain truths about themselves.

[3]Or the idea of the true. Rest assured that I make no claim to having a privileged relationship with truth. I very much like to propose poetic definitions of the "truth". For example: truth is a queen who wins every war without an army. Or even: truth is a goddess that no god could ever completely undress. Truth is something you will never read in any newspaper, not even something you might divine for yourself by comparing two independent newspaper articles, or that you might guess even better by comparing three etc. Truth is the source of doubt: the more you know, the more you don't know; so said Socrates and Jean Gabin. Truth is nothing more than the hope of our conscience.

[4]I will always use the term "communicate" in the sense of honest or scientific affirma-

Just like the amoeba, these widgets seemed to be intrinsically incapable of affirming certain propositions, certain truths concerning themselves.

What truths? The consistent quality of the self. The fact that *one will not communicate the false.*

Here is an honest entity that, due to its honesty, is completely unable to assert that it is honest. Thus, among honest widgets, those who assert that they are honest are by definition, dishonest. Following this realisation, I developed an irresistible attraction to these widgets! I found them amusing and pertinent. This time, it was no longer a question of abstract biology, but frankly of abstract *psychology*, and this psychology concerned incommunicable truths, similar to the amoeba's secret! The most wonderful of all, if I may dare anticipate Nagel & Newman, is that these entities seem able to prove that *if* they are honest, *then* they are incapable of communicating the fact; just as my amoeba seen under the microscope in the act of self-division "told" me that it could not possibly make any claim to having survived. Each of its siblings asserted it implicitly by pointing a pseudopodium at *the other* amoeba! If one of them was another one of them, they could very well be *others*—both of them.

Gödel's theorem and his proof showed me the existence of (self-) reproductive entities; of abstract amoebas as much as the existence of whatever-you-likes incapable of asserting certain self-referential truths such as the consistent nature of self. This was exactly what I had been looking for. The whatever-you-likes in question were formal theories like Peano arithmetic or the *Principia Mathematica* of Russell and Whitehead. No longer in any doubt, I resolved to become a mathematician and to specialise in mathematical logic.

Note that at this time, I was suffering from an immense handicap: I had not yet heard any talk of Church's thesis or of the computer! The term "computer" evoked in me visions of the immense and rigid refrigerator lookalikes used by bankers. I had no idea that a century earlier, Babbage had dreamed of a device calculating the positions of heavenly bodies. I had no inkling of a "Turing Machine". I did not yet truly know what I was up to or what I was conjuring with in terms of informatics, much like Jourdain with prose. Broadly speaking, Alonzo Church's thesis says that:

All entities are machines (Tous les machin sont des machines)

Or better: anything formally calculable (and relatively communicable)

tion. I identify or model, here and further on, this type of communication with a formal (or the formalism of a) proof.

can be calculated (and relatively asserted) by computers (*relative* to a formalised theory).

I still had no idea that machines were actually widgets (the reverse of Church's thesis) or that "universal machines"—computers—were at one and the same time close matches to formal theories and as such, very likely candidates for self-reproducing entities. I only realised all of this much, much later. (In any case, information technology was not yet a study, a subject in its own right at university, merely an option for mathematicians and engineers.)

What, precisely, had I seen in Nagel & Newman? I will use a little freedom in explaining the idea. The basic technical concept appears in the current chapter header: *If DA gives AA; and DB, BB; and DC, CC; what gives DD?* The certain answer is that DD gives DD. Otherwise put, the "replicator" D which makes AA from A; BB from B, ... when applied to D itself gives—as a *result*—"DD" which is to say that the equation describing D when applied to D, is the founding equation itself.

Essentially, if an environment is rich enough to support replicators, then it is rich enough to support self-replicators, the product of replicators acting on themselves.

Another example. Imagine that in a given formal language there occurs an equation describing the operation of substituting an unknown X with a certain formal expression, for ex. $\ulcorner abc \urcorner$ in another formal equation $\ulcorner baX \urcorner$. The quote symbols of $\ulcorner$ and $\urcorner$ are used like inverted commas at the heart of the formal language: they prevent the evaluation of the quoted equation. One uses such substitutions implicitly whenever text is edited by a computer involving "seek/replace"-type operations. In this formal language, such a notion can be set out as follows:

$$subst(\ulcorner abc \urcorner, \ulcorner baX \urcorner)$$

Let us now consider a *widget* capable of interpreting this formal equation, of assessing the result of the described substitution, in this case *baabc*. It is understood that the operation *subst* replaces the 'X' or 'X'es of the second equation by the first.

Indeed, it is the 'X' of the quoted equation appearing on the right that should be replaced by the quoted equation appearing to the left. Thus:

$$subst(\ulcorner aXc \urcorner, \ulcorner baX \urcorner) = baaXc$$

For such a widget, you could convince yourself that the following equation

$$subst(\ulcorner subst(X, X) \urcorner, \ulcorner subst(X, X) \urcorner)$$

is self-referential. It is truly a question of a very simple procedure, with incredible consequences as, hopefully, one may see from what follows.

This procedure, for constructing self-referential equations is referred to as *diagonalisation*. The very term is descended from the fact that if $A(x, y)$ represents an effectively infinite matrix or lookup table of numbers, then $A(x, x)$ represents the diagonal of that lookup table. The construction of self-reproducing entities puts in place *two* diagonalisations, or one diagonalisation applied to itself. In effect, one constructs '$subst(x, x)$' (first diagonalisation), next 'x' is replaced by '$subst(x, x)$' in '$subst(x, x)$' (second diagonalisation).

There follows a useful generalisation of this technique. Imagine that you desired to find an equation which, instead of producing a version of itself, produced the result of a transformation T applied to itself. In the current chapter header, it suffices to designate a new operant D (and I still notate this as D)—which (this time) applied to A gives T applied to AA, which I simply denote $T(AA)$. This obtains no matter what the value given to A in the equation. In this case, D applied to itself, namely DD, gives the result of the T transformation when applied to itself. Similarly, using the substitution subst, it suffices to replace $\ulcorner \mathrm{subst}(X, X) \urcorner$ to gain the formal equation:

$$\mathrm{subst}(T(\ulcorner \mathrm{subst}(X, X) \urcorner), T(\ulcorner \mathrm{subst}(X, X) \urcorner))$$

from which the interpretation will again give T as applied to the formal equation itself. Once again, the consequences will be incredible—as I will shortly demonstrate, even though I make no claim of having instantly and clearly grokked all this at the time of my first reading of Nagel & Newman. Just as Joël de Rosnay was a springboard for Watson, so Nagel & Newman acted as springboard for Kleene's 1952 *Introduction to Metamathematics* [32] and Ladrière's 1957 *Internal Limitations of Formalisms* [35], both of which were alas, out of print. I nevertheless came to extract the two works from the National Library in a veritable act of heroism made possible thanks to a friend whose father worked there. This friend, Dominique, would share with me numerous metaphysical speculations and together we would study our "Ladrière". Today, given that computers literally swarm like lifeforms, you have probably worked out that a computer is effectively an entity of the type described here, capable for example, of correctly producing substitutions, and therefore susceptible to "infection" by a self-reproducing or self-referential equation.

One consequence then appears almost immediately: a computer cannot resolve all questions put to it. In particular, it cannot instantly resolve the

dilemma of knowing if an arbitrarily selected machine will halt or not, once launched on its execution script.

In effect, if such were possible, we would have at our disposal one of these machines—let it be named STOP?—capable, when applied to another machine[5] X, of deciding whether X will halt or not.

But we could well devise a new equation (a new entity) as follows:

if STOP? (X) then CONTINUE else HALT

CONTINUE is an instruction (an equation) that launches the computer into an infinite loop and HALT, on the contrary, halts the computer.

This equation defines a certain transformation T which may be substituted in the self-referential equation of the generalisation described earlier. One obtains:

subst(if STOP?(subst(X, X)) then CONTINUE, else HALT,
if STOP?(subst(X, X)) then CONTINUE, else HALT)

The evaluation of this expression will be difficult to read, but for a computer, it is equivalent to p with:

$p =$ if STOP?(p) then CONTINUE, else HALT

or even

if STOP?(ME) then CONTINUE, else HALT,

which is capable of deciding that it stops (when given to the computer) and in this case to continue, or of deciding that it doesn't halt, in which case it halts. That is absurd, so there cannot be a machine like STOP?. In machine terms, no machine is capable of deciding in a general way whether an arbitrary given machine will halt or not.

Gödel showed in a similar way that with sufficiently rich formal theories (in terms of *mechanically* provable arithmetic propositions), for every predicate[6] $P(x)$, there is a precise proposition q such that the formal theory can show $q \leftrightarrow P(\ulcorner q \urcorner)$. Proposition q is self-referential—it refers to itself. It is an elementary form called the *diagonalisation lemma* in the literature[7]. The proof of this lemma only requires showing that the theory is capable of proving elementary truths concerning substitution. That explains the gigantic reach of this self-reference lemma.

We get, for example, Tarski's theorem–the non-definability of truth–in translating it into the language of Epimenides's paradox:

[5] or to the formal description of that other entity.

[6] A predicate is the formal equivalent of a definable adjective in the theory's language.

[7] The term "lemma" is used by mathematicians to denote a preliminary result.

I am not a true proposition

One can no longer suppress the paradox, as Russell and Whitehead's *Principia Mathematica* does, by banning self-reference, because it is undeniable for systems capable of elementary manipulations such as substitution. The thing that Tarski also proved is that the notion of truth (or a proposition) for a (sufficiently rich and consistent[8]) formal system, is not definable from *within* the system itself[9].

By contrast, Gödel showed that provability by a (sufficiently rich) formal system *is* representable *within* the formal system, that is easily conceived, in light of the notion of formal proof, as an essentially combinatoric notion, in contrast to the notion of truth. Epiminedes's paradox, with "provable" in place of "truth", leads to Gödel's incompleteness theorem of 1931.

In effect, the proposition:

I am not provable by theory T

is representable in theory T, assumed to be consistent, and is therefore true but not provable in T. In effect, if the proposition were false, in light of what it states about itself, it would be provable, and T would prove a false proposition and thus be inconsistent.

I think, therefore, that Gödel's theorem illustrates the existence of a mathematics allowing

1. the revelation of the amoeba's secret without falling into the trap of communicating the incommunicable. The idea is a cross analogy between "I live", "I survived", "I am conscious" and "I am consistent". The self-duplication thought experiment already illustrated the incommunicability of survival. For the amoeba, as for Nagel and Newman's widgets, it seems that there are truths that are quite simply not communicable.

[8] A formal system, or theory, or machine generating propositions, is said to be consistent where it does not prove false propositions, or contradictory propositions like $p\&\neg p$. "$\neg p$" denotes the negation of p. If p is true, $\neg p$ is false, and if p is false, $\neg p$ is true.

[9] More precisely, call a predicate $T(x)$, a truth predicate, if the widget (machine, theory) proves $p \leftrightarrow T(\ulcorner p \urcorner)$, no matter what the value of p is. If $T(x)$ is definable in the theory's language, we could define a falsity predicate $F(x)$ ($F(x)$ is defined by $\neg T(x)$), so that the theory proves $\neg p \leftrightarrow F(\ulcorner p \urcorner)$ no matter what p is. But by applying the diagonalisation lemma to predicate F, we uncover a proposition q such that the theory proves $q \leftrightarrow F(\ulcorner q \urcorner)$ The machine (theory) therefore proves that the false proposition $q \leftrightarrow \neg q$. $T(x)$ is not definable therefore, within the machine's language. This theorem, due to Tarski, plays a role in chapter 8.

2. the offer of a rigorous framework where one could operate the epistemological reversal between biology (or psychology, theology) and chemistry (or physics). The widgets in questions seem furnished of a fundamental mathematical biology and psychology, independent of chemistry. The remaining work illustrates the usage of this mathematics.

To my mind, Gödel declared biology to be more justified than chemistry. Mathematical logic gave me the sense that we can study, in a general way, the discourses of machines (widgets at that time!) obtained from those that self-observe or introspect. It seemed to me that physicists' work is a particular case that must be justified from this more general theory. Watson said that the cell obeys the laws of chemistry. With Gödel's incompleteness theorem, I glimpsed a communicable model of reality where to the contrary, it is chemistry that obeys the law of the cell, where the cell becomes an abstract amoeba, that little entity self-referentially correct to *an*[10] (at the time) universal environment (in the Church-Turing sense). Gödel's theorem largely accentuates the *tangible* character of mathematical reality, and it seemed to me that chemistry could be considered productively to be the product of dreams and coherent discourses of immortal amoebas.

A particular event was significant in this respect. It consisted of a quite animated discussion with my friend Dominique on the fundamental status of different sciences. At this time, Dominique affirmed that physics was the fundamental science. The discussion was bitter because it was about deciding the choice of university studies.

According to me, Physics cannot be the fundamental science. The idea was that we could understand more by understanding how a brain "looking" at the universe produces a theory of the universe, than by understanding the "Theory of the Universe". And if the brain is similar to the amoeba or a formal theory, this process of comprehension does not depend on the nature of the material from which it is constructed. Ultimately, it would be necessary to explain sooner or later where the formal systems come from, aka "widgets", and the belief(s) that a universe exists or that matter exists. Without doubt, because of my realistic childhood dreams, but also because of my fear of believing in non-existent things, I have *never* taken the existence of matter for an established truth. Now, with the appearance of a biology and a psychology independent of the laws of matter, I started to

[10]It took until 1987 for this point to become clear. Correct self-reference would no longer be defined relative to *a* universal environment (universal machine), but relative to the most *probable* or *credible* universal environment.

think that the concept of matter must be explained in more primitive terms; of the beliefs *of* certain "widgets".

Chapter 4

Darker Than You Think [I] (1973→1977)

It is possible to destroy someone with just words, looks, innuendo: this is called perverse violence or moral harassment
Marie-France Hirigoyen

I finally took up candidature in Mathematical sciences at the Université Libre de Bruxelles, with my two suitcases of Biology and Chemistry, and "Gödel's proof" in the hope of being able to link one to the other, perhaps not in the sense given by Watson.

After the first hour of the logic course I asked the professor of logic, X, (so as not to name names), if he was going to cover Gödel's theorem this year, or later, because ...I was about to divulge, somewhat naively, my motivation for Gödel's theorem, telling him that this is the theorem that inspired me to take up mathematics in order to cut the Gordian knot between Biology and Chemistry, etc. But I hadn't the time (I *never* had the time) to finish the sentence. As soon as he processed Gödel's name, he interrupted me abruptly with "Forget Gödel's theorem, there is nothing interesting there, its a finished story".

This was evidently false. But at that time, I didn't know. It is true that my Kleene dated from 1952, and Ladrière from 1957. I took him at his word, and impatient to get into it, followed from the first year, and with his agreement, accompanied by my friend Dominique—who was enrolled in physics, finally—the entirety of the logic course given by this professor in first and second year, at the expense of the other courses. And a certain friendliness settled in. His course on model theory was interesting.

He drove Dominique and me to the Mathematical logic seminar at Louvain, but not to Philosophical logic seminars; even though at Louvain, these seminars brought together philosophical logicians and mathematical logicians. Philosophical logic, nevertheless, is always as mathematical as mathematical logic.

"And intuitionist logic?" I asked him one day. "It's frankly idiotic", he responded. "And modal logic?", "leave that to the philosophers". Etc. The worst is that I took this attitude as a form of humour and that I continued to admire him without really understanding why! It was in part, because I admired a logician who could teach me all sorts of interesting things in logic, all the time showing me a certain form of deprecating humour that I found entertaining.

I got to the second license[1] (fourth and last year) without problems. In between, I again frequented the Molecular Biology Laboratory of ULB at Rhodes-St-Genèse, and even though I spent more time in the library than in the laboratories, I often discussed the "logic" of the lambda phage with René Thomas and his student Jean Richelle[51]. I would study this in depth in my end-of-studies dissertation.[2] Ladrière remembered me and offered me a copy of his wonderful book on Gödel's theorem, and invited me also to Louvain to present the logical and biological work of René Thomas. I also made several evening presentations on Gödel's theorem at ULB, of my own initiative, with the encouragement of my co-students, re-inspired by the 'Ladrière'.

By prudence, I never promised anything to to René Thomas, but I finished by asking X, with my tongue barely in my cheek, however, if it was conceivable that he could supervise my end-of-studies work with the collaboration of Thomas (an interdisciplinary work, all-up).

I hadn't dared tell him that Thomas proposed that I deepen the relationship between systems described by discrete logic equations and systems seemingly described by differential equations. I no longer dared to tell him that I ruminated on a personal project that I wished to suggest to René Thomas. I wanted to see whether Thomas's logic equations, those that he managed to get the bacterium *Eschiria Coli* to execute, were sufficiently rich to calculate recursive functions. In modern terms, that would be coming to view a bacterium as a (little) computer. In any case, as you might have

[1] If one ignores the fact that twice I forgot to do the probability calculus exam, and other anecdotes of that nature.

[2] *Le mémoire de fin d'études*, for which there is no internationally accepted term, but which corresponds to an honours thesis in Australia, dissertation in the UK, senior thesis in the US or major paper in Canada.

guessed, Gödel's theorem applies to bacteria, cells, to amoebae, etc.

I hesitated because I anticipated two trying ordeals. If I opted for the subject suggested by Thomas, I feared the need to come to terms with the world of differential equations. That usually plunged me into an abyss of perplexity that always lead to questions such as what is a real number, what is the continuum, and what about Cantor's Paradise. The other option (secretly linked to the amoeba's secret) risked being a test of knowing whether X's remarks on Gödel were truly lighthearted or not.

I was there in my thoughts, when, without further explanation, X gave me his approval for an end-of-studies work in collaboration with René Thomas. What a guy! I knew he was open-minded! So therefore I had to face the perennial dilemma, the discrete amoeba or the continuous one?

I probably didn't reflect for long, because soon after, I was enrolled in my end-of-studies work with X, and he let me know that he had a change of opinion. "Forcing[3] or Admissible Sets" he suggested to me. And then charged me with summarising an article and giving him a progress report each week.

This was a trying ordeal, that seemed to endure longer than all the previous years. X passed his time in showing me that he was quite as smart as I was an idiot, rasping off any originality that I could slip in, so much so, that my end-of-studies work was nothing but a resumé *by* X, of an article chosen from the literature, with the exception of a small but original section on my part, purely mathematical and technical, that I managed to preserve for better or worse.

Without insulting me explicitly, it became clear from the start of the ordeal, that his aim was to convince me that I was *truly* a "complete idiot", absolutely inept at such an academic career, and he marked "my" end of studies work a 15, effectively preventing me from gaining any of the available diverse national research scholarships. To understand this "15", I asked X if he had found errors in the original part. X, putting on an astonished expression, let me know that in any case, the mark could no longer be changed.

As for international scholarships, a letter of recommendation is required. I could not decently ask one of Thomas, whom I dared not see, since by now I was ashamed and no longer believed in myself.

The amoeba was very distant, and logic made me sick to the core.

[3]The name of a technique invented by Paul Cohen, a student of Gödel, to show the independence of certain formulae in set theory. "Admissible sets" was introduced by Kripke and Platek, and are beyond the scope of this work.

That all happened back in 1977. It was only recently (in 2000) that I understood that I had suffered what today would be called a moral perversion, or even psychological harassment. This is often associated with vampires, and it is true that it takes something from one's life. I don't know what I would have done if I had understood earlier what had happened to me. Am I complaining? Earlier, some considered me paranoid—there are those students who complain when they are badly marked. I didn't complain, I didn't even think to: on the contrary, I felt guilty that I could still be interested in Gödel's theorem, when I was *certain* to be an idiot[4].

Completely demolished, and intellectually perverted, I felt somewhat happy with my lot: having convinced myself that I could not measure up to "real" logicians, and not being able to measure myself against anyone else, X made me almost glad about my failure to pursue an academic career.

[4]A certainty that I came to *doubt*, I'm happy to say. But it would take time and luck, as the following story will illustrate. Note that this wasn't just some *incidental doubt*, but a welcome doubt that I would wish upon everybody.

Chapter 5

Dear Freedom (1977 → 1987)

Take a detour, when the road getting there is closed.
Taoist proverb.

"Goodbye calves, cows, pigs, the whole farm ...". The idea of doing a university degree had been a beautiful dream. For now, I must get my life back, and my sense of freedom and happiness.

Besides, didn't I have every reason to be happy? Happy, first of all, that the ordeal of the end-of-studies dissertation was over. I was planning to do a "normal" pure mathematics thesis, but the idea of prolonging any kind of academic interest evoked the prospect of an ordeal. I clearly thought it would be an ordeal, in light of what I would attribute—more or less consciously at this time—to my incompetence.

Even though I had not let out a peep of my pre-university investigations during my studies, other than "Gödel?", the avalanche of "blows" this single question caused, not only gave me a distaste for Logic, but also for Biology. The Gödellian amoeba was displayed on the placard of my childhood fantasy. But if the Gödellian biologist was dead, the chemist arose again in me. Certainly with some trepidation of the mathematics involved (strange for a mathematician), but assuaged and happy in the end that it would only be a hobby of the mathematics teacher that I would become, for the next six years, and others beyond, in diverse schools and institutes in the city of Brussels.

I was happy therefore, at recovering my freedom of thought, a prerequisite for serious fundamental research.

Finally, happier still that I had preserved my taste for oral presentation, and that I appreciated very much—as I still do—the profession of teaching, particularly of Mathematics.

Happy perhaps, but surely sad. A slight depression slowly overcame me.

The remainder of this "intellectual" story is a little tortuous. I am going to try to summarise several principal events and will return to the quantum chemistry detour.

The awakening of the chemist, in effect, led me rapidly into Quantum Mechanics, and this time I took notice of the truly quantum peculiarities[1]: indeterminism, inseparability, the measure problem, etc. In 1978, I wrote a "mini thesis" on Bell's inequalities, in which I mostly ask questions. Then I followed a more conventional path. The troubling nature of reality illustrated by quantum mechanics, made me search for other world conceptions, as far as possible from the prevailing Aristotelianism. Motivated by taoist philosophers—Lao Tzu, Lieh Tzu, Chuang-Tzu— I started by following the classic Chinese course. I had even read the materialist and immaterialist Hindu and Platonic doctrines.

Then we come to a quite intense period where two conceptions of reality battled it out within me. The war between the mechanist biologist, who was no longer entirely immaterialist—and was probably much less so than during my childhood—and the mystical chemist, ultra materialist, took the form of a confrontation between two interpretations of quantum mechanics: that of Wigner's: *a priori* non-mechanist and quasi-idealist, where consciousness constructs reality in a certain way; and the more mechanist (and a priori more materialist) one of Everett where every possible consciousness is supported by a relative possible reality. I will return to this below. It is in Everett's more mechanistic perspective that I will return to the translator argument (classical teleportation[2]). It consists of an implicit return to the immaterialism of the amoeba. I described in my 1980 diary two possible "experiences":

1. *The minor realisation.* In brief, the understanding of one's own immateriality. In my 1980 diary, I describe that as a possible, quasi-mystical experience that one could have, but I insist that one can deduce the

[1] In contrast to the peculiarities grounding all of physics, which for me was typified by the presence of the continuum.

[2] See the following chapter, or the thesis. The teleporter starts by scanning something at a sufficiently fine level of description, then destroys the object, followed by reconstructing the original elsewhere from the information it obtained during the scanning. Belief that an amoeba survives duplication, belief that one's self survives teleportation, belief that one could survive with an artificial brain or body are, in a manner of speaking, the belief in computationalism. Do not confuse the classical teleportation described here with quantum teleportation. There is a relationship between the two, but it is outside the scope of the current work. I had come up with this notion independently, and used the terms translation and translator.

argument of the translator. We could explain this immateriality to anyone who accepts classical tele-transportation as a means of locomotion. There is also no need to understand quantum mechanics to understand the argument. Obviously, it is connected to the argument that runs: if an amoeba survives a replication, its identity is in its form—not its matter.

2. *The Major Realisation.* In brief, I described it in many ways in the 1980 diary, often in very "mystical" terms, like the disappearance of self or universe, but most often by the "equation" WIGNER = EVERETT. I will return to this subject a little later. This experience is described as being exclusively mystical and non-communicable. As always, it contained the indubitable weakness of my fundamental approaches, and that made me desperate. I would realise later that this type of "mystical" experience was a veritable lure created by my mind to make me accept the idea of the existence of matter, that my "childhood rigour" was thereby seriously shaken.

I had profoundly "regressed" in respect to my intuitions of 1963 and 1971. The chemist (within me) was materialist.

There was, even so, some pedagogical progress. The translator/transporter illustrated the communicable part of one's own immateriality for those who accept its use as a means of locomotion: the idea that I formerly expressed, albeit badly, with the amoeba of my childhood.

Certainly, I never spoke of this at university. I had nevertheless tested the Translator Argument over a year (1980) on my friends at the Café Beppino—to the point of annoying them sometimes! These conversations, and opinion-surveys on the question "would you get in the transporter" were mentioned in detail in my 1980 diary.

Concerning the major realisation, I kept quiet. Once again, I had the feeling of profound truth, but totally incommunicable. I also wondered to what end it might serve, in a world with starvation, of discovering a fish so large that nobody could ever catch it. The problem was that I always repressed the idea of returning to Gödel, *as* a means of communicating the "incommunicable". The situation was, however, more complicated than I thought at that time, in that the *major realisation*, which I had linked to a form of quantum materialist mysticism, actually contradicted the *minor realisation*. I connected mechanism with matter (like everybody else), without taking it into account (as the 1971 Gödellian justification was repressed), all the while knowing that mechanism implied a non-definable immateri-

alism (and without doubt that the "amoeba" was at least as repressed as "Gödel").

Future progress consisted of returning to the childhood intuitions (1963 and 1971) of removing immaterialism from mechanism and Everett, and materialism from Wigner. But I was a long way from this possible move.

And so a certain depression overcame me after this intense period. I became vegetarian, and I took more and more to Zen mediation to end up quasi-completely *immobilised* by the end of 1980.

Alright. I didn't find my 1981 diary. Nor the '82, nor '83. There were several significant events, however—significant for the development of the future "thesis", I remember.

- My friend André bought himself a TRS 80 computer and demonstrated it to me[3]. It was at *that moment* that the penny dropped for me about Church's thesis and I realised the importance of the Universal Machine. In my turn, I bought a TRS 80 and studied its functioning, and increasingly theoretical computer science. I was still nauseous of logic, which in the world of computer science is quite some handicap. I recall that X never once mentioned Church's thesis in his course on computability. This was an omission that artificially quarantined the subject and prevented me from realising the import of Gödel's theorem to the world of digital machines. His course no more mentioned Church than Gödel!

- The appearance of the remarkable book by Judson Webb "Mechanism, Mentalism and Metamathematics" [66] which showed notably how Gödel's theorem is a confirmation of Church's thesis. It developed "my" 1971 intuition, but this time with a clear relationship to the Universal Machine. I will recount this later; I didn't open the book straight away, in effect because I feared discovering that X was wrong (or I was wrong) and of discovering that Gödel's theorem was as alive and well in the field of Mathematical logic as it is in the Philosophy of Science. I didn't open the book because of the nausea I felt for mathematical logic.

- My friend Corinne returned from the USA with a copy of Hofstadter's book "Gödel, Escher, Bach" that she suggested I read as soon as possible. In reality, she suggested making a video on the theme of self-observation or introspection for an art exhibition——this we did. In a

[3]TRS = Tandy Radio Shack

sort of moment of crisis, I read Hofstadter's book, in the countryside—*three times* over!

My first impression was that it is a very beautiful book, and an original introduction to Gödel's theorem. It doesn't extend the idea, however, and, like myself since 1971 (in contrast to Webb), it ignored the Universal Machine, except for a chapter on Church's thesis[4]. In fact, the Universal Machine is the major oversight in Hofstadter's work. Nevertheless, I found pertinent his usage of Gödel's theorem in favour of the possibility of artificial intelligence, as well as his critique of Lucas's argument [5]. In a certain way, this book encouraged me to study Artificial Intelligence (AI). With time, I think that this work has perhaps deterred AI researchers from Gödel's theorem. Hofstadter beats around the bush. A good bush, but he prances too quickly around it, and by a sort of centrifuge effect, he departs, along with his reader, from the idea that Gödel's theorem is truly important for cognitive science; from the idea that it could, for example, be the first theorem of exact psychology,[6] an idea put forward by John Myhill in the 1950s, as I discovered later.

- The appearance of Hofstadter and Dennett: "Minds' I"[20]. For the first time, the book discouraged me because I read there what I could have written best under the subject of the "minor realisation". Unfortunately, that book became a "Watson" and certainly remains the best introduction to my thesis. I recommend it to those who study my work, as I also recommend the remarkable little science fiction book "Simulacron 3" by Daniel Galouye[26]. But neither Hofstadter, nor Dennett made the connection between Gödellian non-provability and the incommunicability of surviving the teleporter, to the point where I continued to doubt the pertinence of the association glimpsed in 1971.

- The discovery of cannabis. In 1980, following the lecture by Alan

[4]It is that which allows the addition of the qualifier "Turing" in front of "Universal Machine". See the following chapter.

[5]Lucas proposed an argument in 1959, according to which Gödel's theorem showed that we are *not* machines. In fact, the argument can be found already in Emil Post's notes of 1921. See the thesis for more information. See the 1995 IRIDIA technical report for a detailed description of the relationship between Mechanism and Gödel's incompleteness theorems.

[6]But Hofstadter, above all, made Gödel's theorem fashionable, and "serious" people want to be fashionable. For example John Haugeland[30] said part of his regret was not introducing this or that subject, but without the least justification, he expressed having *no* regret for not talking about Gödel's theorem!

> Watts, I decided to test cannabis myself. As my parents taught me to be wary of unknown things, I contented myself with planting three seeds and during the time that these plants grew, I firstly read the maximum of literature on the subject, as much "pro" like Solomon Snyder as "anti" like Gabriel Nahas. However the systematic vehemence of the latter indicated to me the innocuousness of the substance. This relative innocuousness, compared to tobacco or alcohol, for example, is today recognised by official health experts in most European countries[7]. It is recognised, however, that cannabis makes bearable certain ailments, such as those provoked by chemotherapy. For me, cannabis made bearable the nausea and disgust I had for logic. It is possible that weed even creates a partial amnesia, shattering connotations that one builds up in the course of life. Surely useful when those connotations are negative. Cannabis allowed me to shortcut meditation, attaining a seeming state of relaxation (let's say) for the first time. The second time would eliminate the negativities altogether, and in any case, made them very rare; this represented a considerable gain in time, but also a relief for my knees![8].

With all that, the Gödelian biologist was "resuscitated", after a certain fashion.

At that time, I found my passion for Artificial Intelligence, I was called up for military service. I therefore opted to do community service as a conscientious objector, and Georges Papy, professor in the Mathematics department of Université Libre de Bruxelles, suggested I do this service in the Algebra department (1982–84). The task was forbidding: pedagogy of Computer Science. I gave courses in Computer Science to all sorts of people: children, handicapped children, teachers, students, professors, etc.

So it was that in 1982 I was invited, thanks to Professor Papy, to perform my community service, at Arlon, Belgium, in teaching computer operation to young people. I made an eight hour presentation, simultaneously translated into Italian, "Computers are graphs", which gave rise to an Italian publication[40] in 1983. This article is the progenitor of the "Movie Graph Paradox", which I described in a 1984 diary, being a means for illustrating the difficulty that the mind–matter problem has with the mechanist thesis. The MGP became the Movie Graph "Argument", and is used in the thesis to eliminate a supplementary hypothesis in the principal demonstration. I published this argument/paradox in 1988[39]. An American, Tim Maudlin,

[7] See, for example, [53]

[8] Lookup "meditator's knee"

published a conceptually-equivalent argument in 1989[43]. Maudlin's proof is more informative than mine. The Movie Graph Argument is presented in chapter 4 of the thesis.

After my time as a conscientious objector at ULB, I gave a course on the functional programming language LISP and on logic programming in PROLOG, as well as an introduction to artificial intelligence[9], lambda calculus, combinators and neural networks.

These courses met with a certain success, with the exception of the computer science staff at ULB, who did not want to hear of artificial intelligence, nor of logic, nor of functional programming.

On the other hand, at VUB, the *Vrij Universiteit van Brussel*, the Dutch speaking version of Université Libre de Bruxelles, Luc Steels had returned from MIT in America[10] with money, projects in AI and LISP machines. So, I worked at VUB, giving several seminars in Flemish on the "introspective capacity" of universal machines and its possible application to Artificial Intelligence. As a result, I was offered a research assistant role in Computer Science. Even though I had perfected my Dutch, which was learnt at school, in an intensive course of four hours per day over a period of three months, they refused me the research assistant position because I still had a pronounced French accent! Luc Steels was not responsible; it was the faculty president who feared that my French accent would pollute the minds of the Flemish students! I had encountered here a sad and well-known Belgian problem.

I mention that I had also worked, before the community service, for a professor of Mathematics and Psychology, Monsieur Ducamp, one of the first, along with an engineer, Pierre van Nypelseer, to be interested in artificial intelligence at ULB. This allowed me to have an account on the university's computer.

After the community service, it was suggested that I work for a private company "Plant Genetic Systems", a Flemish biotechnology company based in the town of Ghent, and which invested in a research unit at ULB: Unité de Conformation des Macromolécule Biologique" (UCMB)[11], directed by the biochemist Soshanna Wodak.

As can be seen, I remained at ULB. Before the community service, I

[9] I appreciate Artificial Intelligence, both connectionist (neuronal) and symbolic. I fear, however, that the expression "Artificial Intelligence" is a bit unfortunate. If you take the meaning of the word "artificial" to be "introduced by humans", then the distinction between natural and artificial is ... *artificial*.

[10] Massachusetts Institute of Technology.

[11] Conformations of Macrobiological Molecules Unit

taught in secondary schools, but I had an office in Professor Englert's Quantum Cosmology sector. I also had good relations with the Psychology department. I would however, be regularly invited to interdisciplinary seminars organised by psychoanalysts to present Elementary Logic and Topology, and later to present my own work. Next, I did my community service at ULB, where my office was next door to X's office! On the rare occasions when he came to his office, he practically never said a word to me. Then UCMB, then much later, IRIDIA. Normal enough, though I needed to interact with other researchers and students so I developed a certain savoir-faire in logic and artificial intelligence, which was a very lively subject area. Although forsaken by the faculty of Science—and by the faculty of Philosophy and Letters—there was a real need created by these novel techniques.

At UCMB, I met Michel Bardinaux, colleague, and then friend, a passionate engineer for the ADA language, who pushed me to develop a PROLOG interpreter—for logic programming. I wrote a PROLOG interpreter in LISP in 15 days, and for a year, Michel and I translated it to ADA and optimised it, before using it to automate elementary reasoning on the structure of proteins. That was a fantastic and very enriching experience for me.

Inspired by the work of Ehud Shapiro[55] on the automatic correction of programs, I started to develop the system ANIMA, which is capable of learning by a technique of self-correction. The program perpetually corrected itself at different levels of self-description. This is the work that launched me more profoundly into theoretical computer science, and forced me to concretely recall mathematical logic, notably the analysis by modal logic of Gödel's theorem and more generally of self- reference. I returned to my my Carrollian pilgrimages and bought, in 1986, the books of Georges Boolos "The Unprovability of Consistency" (1979), and of C. Smoryński "Self-Reference and Modal Logic" (1985). I developed a profound interest for modal logic, and I recovered practically all my interest for Gödellian incompleteness phenomena.

The work at UCMB took time and the topics concerning the "thesis" were always considered a hobby that interfered with my profession.

The history of my thesis could be described as a slow return to the crystal purity of my childhood investigations, where the distinction between the communicable and incommunicable part of the amoeba's secret was clarified by the consequences of Gödellian incompleteness phenomena in theoretical computer science.

However, Quantum Chemistry was going to play an indirect, but capital role. To explain this role, I will briefly return to the years 1977 and 1978.

In the development of the thesis, the passage from (quantum) chemistry

is *logically* a detour, nevertheless very useful to understand and motivate the *result* of my work.

Physics in general and Quantum Mechanics in particular continually played an indirect role in this research. Physics could not be the departure point because the physicist takes for granted that for which I search an explanation: the universe or the appearance of the universe, physical laws or the appearance of physical law. Quantum Mechanics would be for me more a target than a base on which to construct a theory.

In 1977, the Gödellian biologist was dead or seriously wounded. The chemist lived on however, and perhaps the physicist too. It is not the amoeba that divides, it is the *universe* that divides the amoeba, the amoeba does nothing, it no longer exists. Death has been buried.

But what is this universe? Is it truly made *of* something, and if so, what? Matter always seemed more elusive to me than life and consciousness.

I rapidly reread Linus Pauling[47], shortly arriving at Cohen Tannoudji Diu Laloë[15] "Quantum Mechanics", promoted immediately to a new "Watson", rapidly accompanied by the two formidable "d'Espagnats": "Contemporary Physics"[21] and "Conceptual Foundation of Quantum Mechanics"[22]. I re-read closely the adorable books by Louis de Broglie[18], that I had undertaken during the last year of my studies, to focus my attention elsewhere.

I asked a physicist friend what this electron was, that had the appearance of passing through two holes at once.

Ah, if only the amoebas were still there! Perhaps, I thought simply, that the electron, in the manner of the amoeba, self-duplicated and passed through both holes. Yes, but the electron can pass through both holes and merge together afterwards. Ah! But sometimes cells fuse also, such as sperm and eggs—which are not freshwater protozoans; of this I hope you are aware! The analogy between the electron passing through two holes and the amoeba that self-divides (or multiplies[12]) is naïve and straightforwardly lame, but behind it hides a key idea that would take time to emerge from my mind [13].

The physicist friend gave me the references to the article by Einstein, Podolsky and Rosen[24] and that gave me a great surprise.

As this surprise is essential for motivating the development of the thesis, let us return for a moment to the question "understand what?" in the spirit of certain physicists.

[12]It is amusing to able to use both formulations. A profound reason appears later: the difference between discourses in the first and third persons.

[13]Simply, the very *Borgesian* idea that histories or computations self-multiply and merge. This will be made more precise later.

We now follow the discussion I had had in 1972–73 with my friend Dominique: I was not satisfied with the physicists' idea that a set of equations could serve as an explanation. Prediction is not explanation, as put well by René Thom. With the equations, when they are not too complicated, we can predict phenomena. But in truth, the equation doesn't *explain* anything. It compresses, certainly, in a very ingenious way, the description of the physical world, but it does not explain the nature of bodies nor why there is a body to start with, nor why these bodies obey laws, nor from where these laws come.

With the article by Einstein, Podolsky and Rosen, I realised that quantum theory is much stranger than I had believed. It somehow explicitly places the physicist in *en flagrant délit* so to speak, without his having a precise idea of the thing he observes. It makes unavoidable the interpretation problem of quantum mechanics in particular, and of physics in general. It presaged also the work of J. S. Bell in 1964[5]: quantum strangeness could be tested experimentally. "Metaphysical" propositions enter the laboratory to be tested!

Quantum theory, in its usual formulation, describes *two* sorts of evolution of a physical system:

1. Schrödinger's equation. It describes the evolution of a physical system when it is not observed. In short, the equation describes a totally deterministic evolution of a wave, which itself describes the possible results of an observation.

2. Wave function collapse: where a certain value is measured, the wave reduces in a non-deterministic manner. The probability of this or that possible reduction of the wave is given by the square of the wave's amplitude.

For example, the wave associated with a particle whose position has just been measured, is spherical and "diffuses the probability" of finding the particle in all spatial directions. If you subsequently repeat the position measurement, you will find the particle practically anywhere.

Until the EPR paper, this phenomenon was explained by invoking a perturbation due to the measuring apparatus. This idea is natural in light of the fact that laboratory instruments are in general much larger than the observed particle.

What Einstein, Podolsky and Rosen explained was that if Quantum Mechanics is taken seriously, the perturbation cannot be mechanical or physical in the usual sense of these terms. In effect, they said, the way that the

waves are associated with physical "objects" entailed that if two particles interacted, then they are only described by *one* single wave. In effecting a measurement on one of the two particles, you reduce the unique wave describing the two particles and this entails perturbing the other particle instantaneously. No reasonable definition of reality can allow such a form of inseparability, according to Einstein, and so QM is false at worst or at best, gravely incomplete.

With the Einstein, Podolsky, Rosen (EPR) paper, the idea that an equation is an explanation is undermined: the equation must still describe a reality, whatever that is, and the EPR paper clearly illustrates that with Schrödinger's equation, this is far from evident.

In 1964, Bell showed that one could experimentally test the existence of inseparability, which led to a series of experiments culminating in Alain Aspect's in Paris 1981[4]. They confirm quantum mechanics, contrary to Einstein et al., but in a certain sense, they confirm the quantum mechanical prediction that Einstein et al. pinpointed: of the existence of "perturbation at a distance". In today's terminology, one says that the two particles are *entangled*.

It is often added that this instantaneous perturbation is random such that it does not allow the instantaneous transmission of information. That is true. Unfortunately, this "reasoning" is added by many in deducing that entangled states cannot have applications. We know today that this simply is not the case. From the preliminary work of Feynman[25], then that of David Deutsch in 1985[23], we know that it is possible to exploit and generate entangled particle states, as with a quantum computer, or with the phenomenon of quantum teleportation. But that will concern us a little later.

There exists no unanimity amongst physicists on the way to interpret quantum mechanics. We can distinguish two families.

- Those who think that after measurement, there is a "real" wave collapse. There is a wave, and its observation provokes a real physical collapse of this wave. For example, if the wave of an electron passes through two holes, the observation of a hole transforms the electron wave into the wave of an electron passing through one hole.

- Those who think such a collapse doesn't exist. Quantum theory is reduced to Schrödinger's equation, *that the observer also obeys*. The observer's wave is entangled with the electron's wave creating a wave describing two observers observing each electron passing through its hole.

The first family of interpretations has many difficulties combining the deterministic evolution dictated by Schrödinger's equation with random state-reduction. The solution proposed by von Neumann and Wigner consists of attributing a special role to consciousness. Physical objects obey Schrödinger's equation; consciousness collapses the wave. No wonder the psychologist Jung (via Pauli) was interested.

The second family requires an explanation of the wave function collapse. That is what Everett started doing in 1957. He showed that if you apply Schrödinger's equation, not to an isolated physical system, but to a coupled system consisting of the observed system and the observer and considered as a memory-machine, Schrödinger's equation only predicts a state-reduction in the discourse of the observer machines in their account of their experiences. The advantage is in justifying the appearance of "perturbation at a distance" and the appearance of indeterminism in a globally local and determinist context. This type of approach will be considerably extended in my work, and even generalised to the truth of Arithmetic in its entirety.

Some more information can be found in the Quantum Mechanics appendix of the thesis.

Schrödinger's equation applied to the observer/thing observed pair predicts that the quantum state of the observer is entangled with the quantum state of the thing observed. If the electron wave described the electron's position as it passes through the two holes, the observer's wave entangles with the electron's wave with the result that the global wave describes a state with the observer seeing an electron passing through one hole, *and* the observer seeing an electron passing through the other hole. The observer has been multiplied in observing the electron. There is no wave collapse, even if it appears so from each observer's viewpoint.

That observers multiply on the tree of possibilities would please the Gödellian biologist, the amateur Lewis Carroll, and Borges[14]. But in 1977–78, the amoeba was forgotten, and it was only slowly that I appreciated the profound relationship between "Turing Mechanism" (Computationalism) and Everett's formulation of quantum mechanics. At this time, contrary to my childhood inclinations, I believed that matter could only be completely mysterious. It was only with the return of my interest in Gödel and the discovery of Church's thesis (see the following chapter), that I recalled that machines and numbers have sufficient mysteries within them for there to be no reason to add more.

Note. I finally found my 1981–1986 diaries. That allowed me to better under-

[14] See his "Garden of the forking paths" in [10]

stand the evolution of my ideas. In short, as I already said, the Gödellian biologist was stunned (let's say). It is clear that it was Gödel's theorem that, in 1971, made me glimpse the possibility of communicating (proving) The Reversal and the fact that Physics and Chemistry could ultimately have an immaterialist basis in possible dreams of abstract amoebas. On learning that Gödel's theorem was out of fashion with logicians (X's "error"), it must be that I unconsciously abandoned this conception of things. Towards the end of my undergraduate degree, and afterwards, however, the materialist chemist within me came back. This time hesitating between two interpretations of quantum mechanics. One on the part of von Neumann or Wigner, quasi-dualist; where consciousness acts on matter to collapse the quantum wave. The quantum wave describes multiple, incompatible histories, with conscious observation "choosing" a history from those described by the wave (cf the electron that passes through two holes). The other, that of Everett, where there is no collapse of the wave at all, and where all histories are "physically realised". Observers, in Everett's interpretation, could be considered as machines with memories. The feeling that they have of the uniqueness of their own history comes down to the fact that *they themselves* obey Schrödinger's equation, *they are themselves multiplied*. I conceived of both interpretations as being materialist (in a weak sense I will elaborate on): with von Neumann-Wigner, it is dualism, or double materialism, a *substantive* consciousness that does not obey Schrödinger's equation, and a substantive matter which does. This interpretation raises enormous conceptual and technical difficulties and I abandoned it for Everett's conception of things. I accepted the Mechanist hypothesis. Forever influenced by the chemist within, I still conceived of it in a Materialist way. There must be therefore, a kind of super universe into which our material universe multiplies. This is, otherwise considered, the usual interpretation of quantum mechanical formalism according to Everett. In 1984, I discovered the Movie Graph Paradox. As for the RE paradox (Universal Dovetailer paradox), I don't have a date. I saw these paradoxes as serious difficulties for Mechanism. And, staying materialist, I saw arguments in favour of dualism almost like Wigner's. Only by the 1986 diary did I realise that the RE paradox, where every computational history is generated, itself generalised the explosion of universe-histories *à la* Everett's interpretation. With the solutions to Schrödinger's equation being computable, this then, is a well defined mathematical generalisation: Everett's multiple histories are a particular case of the multiple, immaterial computational histories. I realised that the quantum peculiarities could very well be a confirmation of Mechanism, because the mechanist peculiarities evident with the movie graph and with the RE paradox are no stranger than quantum peculiarities, as well as being very similar in nature. I no longer interpreted the movie graph paradox and the RE paradox as a possible refutation of Mechanism, but rather as an argument for The Reversal, returning to my intuition of 1963. In the meantime, my interest in Mechanism and finally for Gödel returned (without nausea, thank you, cannabis), and I concluded that the discourse of a self-referentially correct Universal Machine must converge on the combined discourse of Physics and Chemistry, thus returning also to my intuition of 1971. I modelled scientific communication between machines by formal provability, conforming to my "intention"

of 1971. I also put some time into thinking about the use of Plato's Theaetetus theories of knowledge to formally capture the notions of first and third person, as well as modelling the Universal Dovetailer by Σ_1 formulae (see the chapter on the reversal, and the chapter on the machine and its guardian angel).

I realised above all, in flicking through the diaries, that I had abandoned the *reversal* because I preferred to believe in being mistaken rather than believing the universe had tricked me! Later, I would tell myself that Gödel, particularly at Princeton in the company of Einstein, surely must have thought of this reversal, and that, since he never exploited it, must have known that it did not work. Even now, I find it a little astonishing that Gödel and Einstein together had not discovered an Everettian interpretation of arithmetic, that is to say, exactly as rendered logically necessary by this work, as soon as one postulates Computationalism. It is true that Gödel did not appreciate greatly Church's thesis (see next chapter), nor Mechanism.

I realise that there were many other facts I haven't mentioned. A proof that the biologist returned more rapidly than the Gödellian biologist is given by the fact that I studied *planaria* between 1982 and 1985. These are small freshwater worms that are veritable champions of cellular regeneration. They inspired me to my theoretical and Gödellian approach to cellular regeneration and differentiation. I recall my 1992 article "Amoeba, Planaria and Dreaming Machines". Planaria played an important role in my reflections on theoretical biology. A good book on invertebrates was written by Ralph Buchsbaum[11]. Followed by numerous corrected and enlarged editions, it contains a detailed chapter on cellular regeneration.

Chapter 6

The Universal Machine Returns to Earth

Monsieur Bamberger, principal of the Athénée Maïmonide[1], could not believe his ears. For half an hour, yours truly gesticulated in all directions trying to convince him to raise funds to buy computers for the school, for use by the students. I had just invited the most interested of them to come to my place to learn recursive programming with turtle graphics and lists in the language LOGO that I had implemented in BASIC on my TRS 80.

The school wasn't wealthy, the building was barely habitable, and in eternal need of repair, necessitating ongoing costs. I was not too hopeful for these funds.

"And what would the students do with the computers?", the prefect asked me.

"But, sir, think about it! It is a dynamic mirror that will stimulate the grey cells of our students, it is a universal accelerator, a tunnel into other worlds, an epistemological black hole. It is the philosophy machine that you can question, it's the machine that you can start with a verb, it's the Golem, sir, perhaps even more!" I let myself go, because I felt this was in private.

"Monsieur Marchal, I appreciate your enthusiasm, although I think that your suggestions are a little exaggerated. But, as you know, the school is not wealthy, the building is barely habitable, and in eternal need of repair, necessitating ongoing costs. We cannot stop to hold a fête. Also, if you like, we can talk about this some other time as I must go."

Next, and by one of those coincidences that could never happen by chance alone, M. Bamberger went to Israel—to visit schools piloting com-

[1] a Jewish private school in Brussels

puters.

On his return, he met me in his office, and gave me *carte blanche* to introduce computers into the school. I was to give a computer course to staff and students. I organised a computer club, and with a number of sympathetic people, organised the fancy-fair and along with bequests of generous parents enabling the purchase of five magnificent APPLE IIs. At Maïmonide, M. Bamberger considered me a prophet.

Prophet? Me? I do not know if I am a prophet, but if I were the universal machine prophet, I would be such a clumsy prophet. I hadn't recognised the beast, having mistaken the memory modules used exclusively by bankers for stupid fridges.

Above all, I was too late.

I should have been born in the nineteenth century to herald Babbage's machine. That, conceived, and partially constructed in England, was a Universal Machine made of cogs and metal valves! After that, the machine would go to the London museum, where it can be seen working still. Jacques Lafitte's little visionary book "Réflexion sur la science des machines"[36] stated that Babbage suffered more from the lack of understanding by his contemporaries than for his invention of a system of functional notation for his machine. This notation helped him describe its functions and, since he presented the universality of his machine via his system of notation, he must have realised their computational equivalence. Effectively, as I will explain, he presented Church's thesis.

Church's thesis affirms that all possible computers, be they material or virtual, or just as programming languages, are in effect equivalent entities. Ignoring the issue of execution time, they are equivalent in the sense that they are able to emulate (ie simulate perfectly) each other. Together, they define the collection of computable functions. I'll return to this later.

Had I been the universal machine prophet, I would have heralded Turing's Universal Machine. It appeared in the twentieth century as a result of reflections on the fundamentals of Mathematics following the Cantorian mathematical crisis at the start of that century.[2] It is the basis of computer science, and Turing expounded Church's thesis in his 1936 paper[63] where he defined and showed for the first time, the existence of a Universal Machine: a Machine capable of imitating all other machines. Turing was a veritable hero of the Second World War. Unfortunately his varied and important work, ranging from theoretical chemistry to mathematical logic and theoretical and practical computer science; in areas of artificial intelligence,

[2]I have expounded on this point in "Conscience et Mécanisme"[41].

neural networks and the fundamentals of quantum mechanics—was hardly recognised in his own lifetime. He was imprisoned for his homosexuality, and ended by taking his own life.

This brings us back to a moment some 15 years prior, when I should have heralded Post's normal systems. In 1921, in an astonishing anticipation, Emil Post, from the United States, invented or discovered (according to your taste), a universal symbolic system, or as we say today, a Universal Formal System. From a local reflection on finite manipulation methods, he stated "Church's thesis" 20 years before all the others (Turing, Kleene, Markov) in the form of a law of mind. He derived (non-constructively) from this 'law of mind' a general form of Gödel's incompleteness theorem (10 years before Gödel[3]). He discovered the argument, based on "Gödel's" theorem, showing that Man is superior to machine (38 years before Lucas, 68 years before Penrose). He also discovered the *error* in this argument (59 years before Webb, before Bencerraf and many others...). Post is above all known by logicians for having promoted, in his brilliant 1944 article[50][4], a famous problem that would be independently resolved in the USA and the USSR, and that would become the foundation of recursion theory. It is truly a theory of diagonalisation, in fact. It is also the source of inspiration for a very large part of theoretical computer science. In his 1921 notes, he considered the idea of immaterial monism[5] In fact, at that place however, in a note at the bottom of the page, he affirms a change of opinion, and returned in 1924 to dualism, possibly influenced by Turing.

On the other hand, perhaps I should have heralded Markov's algorithms? We can show that a function is computable by a Markov algorithm if it is computable by a Turing machine. And Markov in the USSR, independently of Turing and Post, propounded "Church's thesis".

Always supposing that I'm the prophet of the Universal Machine, perhaps I should have announced Curry's combinators, Church's lambda functions and finally von Neumann's concrete computer, including all those programming languages that have since appeared; languages which define all

[3]Gödellian incompleteness is a more or less direct consequence of Church's thesis. See the appendix in my thesis [42] for a detailed explanation of this.

[4]This article, like those of Gödel on incompleteness, and those of Kleene, Church and Rosser are reprinted in the Super-Watson that is the selection of articles by Martin Davis[17]. The anticipation in the 1920s by Post is in Davis's book, and nowhere else, to my knowledge.

[5]The present thesis shows that Computationalism entails Immaterial Monism, namely the idealist doctrine that matter emerges from mind, not the reverse. Mind, here, is defined exclusively by mathematical truth, or even only by simple arithmetic, comprising both provable and unprovable relative self-referential truths.

those virtual machines and that are above all, universal? After all, the Universal Machine prophet announces these—and not necessarily only those that take into account the implications of universality, either through espousing Church's thesis, or via an equivalent thesis.

OK, but also, perhaps I should have heralded the invention of the telephone by the amoeba. Sorry—I mean to the appearance of the biological nervous system in animals. After all, the brain, which allows us to dream of universal machines, even to construct them today, is assuredly Universal, *at least*[6]. It is perhaps not even difficult for you to to convince yourself, dear reader, that you are capable of emulating a computer, given enough time and space; perhaps then, I should be heralding *you*?

And why shouldn't I announce the appearance of molecular genetic regulation circuits that, manifestly, were able to sustain the emergence of the brain and yourself. It seems that if I were the Universal Machine prophet, I would *have* to herald its eternal return.

In reality, it is the Church thesis that makes the *Turing* universal machine, a Universal Machine—period. And throughout human history, Church's thesis bears witness not only to the (re-)appearance of the Universal Machine, but above all directs our attention to its universality, as much as to its absolute epistemological character on the notion of computability by finite describable procedures.

Curiously, Church himself did not propose *Church's thesis*. He simply proposed to define the notion of 'computable function' by the formal notion of lambda calculable function. It is unnecessary to understand exactly what that means to understand the historical events. It suffices to understand that Church, like Turing and the others, proposed a formal definition of computability. In effect, as Church proposed his definition, Kleene didn't believe it. It is a little absurd to not believe in a *definition*, evidently. Let's just say that Kleene didn't believe that Church's definition was adequate. He didn't believe that it was possible to give both a formal and an absolute definition of the notion of computation. Kleene knew well Gödel's result that showed that the notion of provability is relative. In effect, Gödel had shown, as I sketched above, that the set of provable true propositions in a formal theory is not closed under diagonalisation. In other words, with the diagonal, one can demonstrate true propositions, expressible within the formal theory, yet not provable *within* said formal theory. Stephen Cole Kleene was

[6]This is true independent of Computationalism. Computationalism is the hypothesis that we are no more than a Universal Machine, in the sense where we suppose that a universal machine suffices to emulate us.

persuaded that every notion of formal computability, in particular Church's lambda formalism, must also submit to the yoke of diagonalisation.

Kleene believed he could criticise Church's definition by producing, via diagonalisation, an intuitively computable function that wasn't computable in Church's system, thereby refuting the claim of universality in Church's formalism. Kleene's reasoning went like this: since the system is formal, it entails that each definition of a particular formal computable function in this system be represented by a well-formed formal expression. We could therefore "mechanically" decide if such an expression from Church's system represented a computable function. But then, one can methodically list all computable functions definable in Church's system. It could be done by arranging them by their length (defined by the number of signs in the formal expression), and then arranging expressions of the same length by their alphabetical order. We end up with a list of all computable lambda functions, that, accepting Church's definition, must give *all* computable functions:

$$f_0, f_1, f_2, f_3, f_4, f_5, f_6, f_7, \ldots$$

Now, let us consider the function g defined by means of a *first diagonalisation*:

$$g(n) = f_n(n) + 1$$

The function g is clearly *intuitively, mechanically,* computable. To calculate g applied to n, it suffices to look up the nth function f_n in the list, which is itself mechanically constructed, applying it to n then adding 1.

But the function g cannot be defined by a lambda expression in Church's formal system! In effect, if that were the case, it would be in the list. Therefore, there would be a number k such that $g = f_k$, and a second diagonalisation could be put into play. Specifically, if g is applied to its own number k in the list, we get $g(k) = f_k(k)$. However, by g's definition, we also have

$$g(k) = f_k(k) + 1$$

Substituting $g(k) = f_k(k)$, we get

$$g(k) = g(k) + 1$$

Then subtracting $g(k)$ from both sides of the equation, one gets

$$0 = 1$$

Is not that a formidable demonstration by absurdity, of the incompleteness of Church's formal system, and even of all formal systems claiming to define all computable functions?

Well, no! It so happens that we are able show that the function g is perfectly well definable in Church's system. So what happened when we computed $g(k)$, ie g applied to its own position in the list? We found that $g(k)$ wasn't defined. In terms of execution by a machine, we got an infinite execution, and there is nothing bizarre in infinity equalling infinity plus one.

Kleene's demonstration showed only that it is not possible to mechanically generate the list of all computable functions that are well defined on all their arguments. As soon as you allow functions that are only defined for some values, nothing then prevents one from thinking that the list contains *all* computable functions, as it contains all functions that are well defined for all their arguments, and the diagonalisation does not generate further contradictions. Kleene created the term "Church's thesis", noting that the class of computable functions, not necessarily defined everywhere, is closed under diagonalisation. It is child's play, if you are well versed in Kleene's reasoning, to obtain Gödel's limitation result, and many others, from Church's thesis[7]. For example, it is obvious one can never construct a machine capable of deciding whether a particular function is defined everywhere, from its formal description alone. Specifically, if one could create such a machine, you could mechanically extract from the afore-mentioned list a sublist of computable functions *defined everywhere*, and with Church's thesis, you would have all of them. And this time, with this sublist, Kleene's reasoning would truly show that 0=1. The prize of generality promised by Church's thesis is given with a set of results on the limitations of Universal Machines. The universality makes them unpredictable, essentially uncontrollable. Truly said about these machines, the more one studies them, the more one realises how much cannot be known about them. Church's thesis protects machines from all reductionist theories (completely) that one could invent on the subject. With, *additionally*, the *Computationalist hypothesis*, Church's thesis protects *us* from normative psychologies; it pins the unknown to our very own breast. This will be made precise in chapter 8.

Gödel? Gödel never proposed Church's thesis. He didn't believe in it, for a while. According to what he said, it wasn't until after closely reading Turing's article that he started to accept Church's thesis[8] In 1946, at Princeton[29], he opined that the closure of the class of computable functions under diagonalisation is a sort of miracle:

[7]In this sense, Gödel's theorem *confirms* Church's thesis. Judson Webb, on this subject, said that Gödel's theorem is the *guardian angel* of Church's thesis, and of Mechanism

[8]He never really believed in computationalism. See the IRIDIA 1995 technical report[41] for more information on this subject.

> Tarski[9] has stressed in his lecture (and I think justly) the great importance of the concept of general recursion (or Turing's computability). It seems to me that this importance is largely due to the fact that with this concept one has for the first time succeeded in giving an absolute definition for an interesting epistemological notion, i.e. one not depending on the formalism chosen. In all other cases treated previously, such as demonstrability or definability, one has been able to define them only relative to a given language, and for each individual language it is clear that the one obtained is not the one sought. For the concept of computability however, although merely a special kind of demonstrability or decidability, the situation is different. By a kind of miracle it is not necessary to distinguish orders[10], and the diagonal procedure does not lead outside the defined notion.

Gödel hoped in vain for a similar "miracle" for the notion of provability. But with his own Incompleteness theorem allied to Church's, we can expect that to be difficult. The notion of formal provability is essentially relative, compared to the notion of computability, which is absolute—due to Church's thesis. We will see, however, when we interview the Machine and its guardian angel (in two chapters), how we can formalise indirectly an intrinsically non-formalisable notion of proof (!), this being quasi-absolute, from the *point of view* of the Machine. But that is for later.

It is Church's thesis which allows numerous theoretical computer science results to be *machine independent*, results which do not depend on the choice of formal system used. In theoretical computer science, the choice of machine defines a sort of basis in which machines are identified with numbers. In the

[9] Tarski insisted in his presentation (and I believe he had reason to do so) on the huge importance of the concept of general recursion (or of Turing computability). It seems to me that this importance is largely due to the fact that with this concept one has for the first time succeeded in giving an absolute definition for an interesting epistemological notion, ie a notion that doesn't depend on the formalism chosen. All the other previously treated cases, such as provability and definability, could only be defined relative to a given language, and for each individual language it is clear that the obtained notion is not the one that is being sought. For the concept of computability however, although but a type of provability or decidability, the situation is different. By a sort of miracle, it is not necessary to distinguish orders, and the diagonalisation procedure does not lead outside the defined notion.

[10] This contrasts with formal provability, necessarily relative as a consequence of Gödel's Incompleteness theorem and which can be extended into a hierarchy that logicians often qualify by orders or types.

same fashion as with geometry, where the important theorems are those that do not depend on the choice of coordinate system, Church's thesis ensures that the results do not depend on machine-choice.

In my work, Church's thesis guarantees the generality of the Universal Dovetailer (UD). This is a program, that not only is *capable* of emulating all digital (numerical) machines, but does in fact emulate all of them. One way of picturing the UD is as a *crushed or squashed* universal machine, from which all possible executions flow. It is not difficult to transform a Universal Machine into a Universal Dovetailer.

The only small technical difficulty stems from the existence of non-terminating programs of certain digital machines (a consequence, of the closure under diagonalisation of the collection of computable functions, as Kleene's reasoning illustrates). To construct a Universal Dovetailer, it suffices to construct a generator of all acceptable programs for a given Universal Machine then to zigzag (or *dovetail*) closely on all finite portions over the execution of this machine. It is a well-known technique in theoretical computer science, and depends on the fact that the Cartesian product of two mechanically generated sets is also mechanically realisable[11].

The UD greatly generalises the Library of Babel which contains all books, because it not only generates all books, of which there are an infinite number,[12] but also with Computationalism, all possible readers of these books, and all dreams that these readers might have.

[11]The accepted term, in accepting Church's thesis, for "mechanically realisable" is "recursively enumerable", where the name "RE paradox" for the "Universal Dovetailer Paradox" finally became the "Universal Dovetailer Argument". "Dovetailer" comes from the accepted term in theoretical computer science for this *zigzagging*: dovetailing, which is itself a term used by roofers to describe a means of placing tiles on a roof.

[12]The UD therefore generates all real numbers. There is no contradiction with Cantor's theorem that says the set of real numbers is not countable, ie one cannot arrange all real numbers into a list. The UD does not generate such a list. It generates the reals bit-by-bit, without ever suggesting such a list. The following algorithm also generates each real number in the same way as the UD: generate 0.0 and 0.1, then 0.00 and 0.01 and 0.10 and 0.11, then the 8 after, then the 16 after that, and so on.

Chapter 7

The Reversal ("1963" reprise)

I knew it was always going to be a long explanation of the journey leading to my thesis, with all its detours, and which, as I said earlier, is but a slow return to the crystal purity of the 1963 exposé on the amoeba itself.

One *true* advance was the discovery of Church's thesis and the Universal Machine (see the preceding chapter). A *false* advance was without doubt what follows. The 1963 "result" is "If an amoeba lives two days, it lives forever". I therefore departed on an obsessive quest to prove that an amoeba lives for two days (ie survives its duplication). Now, I have completely (re)integrated the fact that the answer to this question is *truly* incommunicable *in the third* person. Or, if you prefer, scientifically incommunicable. My father, and Ames & Wyler had good reason to remain silent, just like the amoeba, like all self-referentially correct machines, silently demurring on this question. But—the amoeba can place a bet.

Computationalism is a hypothesis: a hope or a belief, according to taste. The beauty is that Computationalism justifies its incommunicable character. This is intuitively justified by the thought experiment of self-duplication, and formally made clear by interviews with the Universal Machine and its guardian angel (see the following chapter).

The question is not, therefore, knowing whether digital or numerical *mechanism* is true or false. The question is knowing *whether you would accept an artificial digital brain transplant.* The question has a relatively urgent character, because we can start by substituting brain parts by electronic artefacts. In particular, a blind person recovers a "sort of vision" thanks to such a substitution. Spectacular progress has also been realised

with animals. The question is not whether one can *truly* know if you are going to survive with an artificial brain, but only of understanding that if you effectively do survive, the reversal of psychology and physics necessarily follows.

It is natural that the logical consequences of incommunicable propositions are themselves incommunicable. I will show in the following chapter how the return to Gödel, and my 1971 intuition,[1] allows us to mathematically isolate the communicable parts from the incommunicable parts in the discourse of a universal machine. By communicable, I mean scientifically communicable, or more simply as I'm going to illustrate in detail, communicable *in the third person*, in contrast to another form of "communication" that I will call *communication in the first person*. The distinction between first and third person, which will be explained here, is another advance, conceptual and pedagogical, to explain more easily The Reversal.[2]

I am motivated by the most recent version of the Universal Dovetailer Argument (see chapter 3 of the thesis), which I presented in April 2000, in Dubrovnik, at the 26$^{\text{th}}$ Congrès International de Philosophie des Sciences.

Remark. The definition of Computationalism presupposes a minimum of *folk* psychology, or *grandmother* psychology, as I sometimes call it. It is the psychology of everyday life. In particular, it is necessary to know how to give a sense to the word "survive" in a minimum of given situations.

When orally presenting the reasoning, I happen to illustrate the minimum necessary grandmother psychology by starting with the following concrete experience. I ask the audience if they will allow me to drop my pencil onto the desk. In general, the audience, a little surprised, allows me. I let the pencil go, and of course, each time it falls. No controversy. I then ask the audience explicitly if they think they've *survived* this experience, and if they would survive if I repeated dropping the pencil. This just illustrates that we are able to give a common meaning to the word "survive" in the sense where we admit to surviving 1001 everyday events. Eventually, step by step, I ask them if they think it possible to survive an artificial heart transplant, and so on, until its an artificial brain transplant. By *definition*, a computationalist is someone who answers yes to *all* these questions.

Concerning the passage from the heart to the brain, I often hear an

[1] Intuition along the lines of Gödel's incompleteness proof illustrates how to separate the provable from the unprovable for classes of formal conversations.

[2] In the following chapter, third person communication is going to be modelled (or even "captured") by formal provability (arithmetisable, Gödellian). Nuances of the type "first person" and "third person" will be captured by the various modalities of Gödellian provability, inspired by Plato's Theaetetus.

objection: "I imagine that I can survive with the heart of another person, but if I get transplanted into my skull the brain of another person, it is rather this other person who survives with my body. The brain and the heart have asymmetric role in this regard". This objection illustrates two things. Firstly, that the person making this objection has a good intuition of the meaning of "to survive". The remark is correct, resulting effectively in the neurophysiologist's hypothesis of the brain being the organ of memory and consciousness. It also illustrates that the artificial brain transplant must be made at the right level.

If you replace chapter 4 of "In Search of Lost Time" by Marcel Proust with chapter 4 of "Alice in Wonderland", the book's contents is perturbed. But if you make the replacement letter by letter respecting the proximity relations between letters, the contents of the books will be invariant for the substitution. It is the same with Computationalism and the artificial brain transplant, the substitution must be done at the right level. Belief in Computationalism is a belief that this level exists.

We will see that this level cannot be determined with certainty, but it can be wagered correctly, as we will suppose in the *gedanken* experiments.

Although it is necessary to admit a minimum of folk psychology, this will be eliminated ultimately to permit a purely mathematical extraction of physics from machine psychology. This is made possible by the Gödellian path, that of 1971, as I will explain briefly in the following chapter. The idea consists of substituting the discourse of the grandmother by the Gödellian discourse of the self-referentially correct machine.

The precise hypothesis of Computationalism is given by the following three sub-hypotheses:

- *The Mechanist Bet.* There is a description-level of myself such that I survive a functional and digitally describable substitution of my components at that level. I call such a level *a substitution level*, or more simply *the correct level*. Another way of putting it: I can survive with a body that is 100% artificial or virtual, ie emulated by a computer. Emulated signifies here: simulated at a level, correct by definition, of substitution.

- *Church thesis.* A modern version is that all computers or universal systems can emulate each other. This was the point of the preceding chapter.

- *Arithmetical realism.* The propositions of arithmetic are true, independently of me. It is belief in the archaic mathematical reality of

which Alan Connes speaks so judiciously[16][3].

So as to facilitate the proof, I will introduce explicitly four supplementary hypotheses, which will then be replaced in a stroke by one new hypothesis. I will show briefly, in referring to my work, how to eliminate this last remaining supplementary hypothesis. I proceed in such a way as to separate out the difficulties. The four supplementary hypotheses are the following:

- CORRECT LEVEL: in the following thought (*gedanken*) experiments, I will always suppose that the descriptions of the body or the brain are done at The Correct Level. This level exists by hypothesis, but does not specifiy that a machine could scientifically determine "the correct substitution level". The computationalist machine can however *bet or wager* on nominating this level, and we can reason in the case where this level has been correctly nominated.

- CONCRETE UNIVERSE: I suppose that there IS a concrete universe, and anything will do. This hypothesis gives a décor for the argument. It is important to see how it might be eliminated. We could speak of "grandmother physics": "concrete" signifies existing in a singular fashion, as everyday objects are supposed to exist.

- NEURO: this the neurophysiologist's hypothesis. I suppose that the level of description of my brain is "sufficiently high". We will see that the reasoning does not depend *in fine* on the choice of level, not even on what we precisely mean by "the brain". The reasoning does not depend on the question, fiercely debated by Anglo-Saxon philosophers of mind, of knowing whether it be necessary to include the environment within the simulation for me to survive the substitution.

- 3-LOCALITY: this is an extremely weak hypothesis that I mention because of the key role it plays in the reasoning. This hypothesis says that, for example, in our concrete universe, it is possible to separate

[3]The position which opposes mathematical realism the most is the conventionalist position: mathematical propositions are purely conventional. This position makes unintelligible the behaviour of mathematicians who hide mathematical results that don't please them. The most famous example of that concerns the Pythagoreans who hid the proof of the irrationality of the square root of 2! Why hide conventions? Realist mathematics has been quasi-unanimously accepted by mathematicians from time immemorial. Most philosophical critiques against realist mathematics are the work of philosophers confounding mathematical theories (necessarily having a number of convention choices) with their referents.

> two computers in a way that means their calculations do not interfere. Those who use a computer implicitly assert, albeit unconsciously, this hypothesis. In the course of the reasoning, you will know why I speak of "3-locality".

The Reversal is a nearly-direct consequence of a preliminary result that I called the generalised invariance lemma (a lemma is the usual word used by mathematicians to denote a preliminary result). The Invariance Lemma states that subjective first person experiences are invariant under a series of objective third person transformations. These terms, however, are going to be defined in such a way as to allow the reasoning to proceed.

The Simple Teleporter Case. With Computationalism and the supplementary hypotheses, the computationalist *practically* survives teleportation (he climbs into the translator or teleporter or transporter or ...). He allows himself to be scanned in Brussels, at a certain level of description of his body. He is then destroyed (all under anaesthesia of course, and in a booth I will call the Scanning-Annihilation Booth) and knowing that the numerical information so-obtained is then sent to Marseille (say). At Marseille, on receipt of this information, the candidate is reconstituted in a Reconstitution Booth. Survival of simple teleportation follows therefore from the fact that there must be transplant into an artificially-engendered body, and that survival depends only on the adequacy of the level of description, not on the method used to reconstruct the body.

From the perspective of an external observer, which I will call the *third person* perspective, the candidate seems to have travelled from Brussels to Marseille. From the candidate's own perspective, which I will call the *first person* perspective, it also seems to be a trip from Brussels to Marseille. In the case of simple teleportation, the distinction between first and third person is fuzzy. That will not be the case in the following thought experiment.

The case of teleportation and self-reproduction. Let us consider the case of teleportation with delay. At Marseille, this time, instead of reconstituting the candidate at the moment when the information arrives, we store it for one year. Then we do the reconstitution. Reconstitution is always supposed to be done in a booth which has no means of measuring time — no windows to the outside world, for instance. But the booth has a self-localisation system by satellite (GPS type), so therefore the candidate can know he is in Marseille. Could he distinguish this experience from the former without delay? With our hypotheses, surely not. From the candidate's perspective, these two experiences are not distinguishable. Let us then define more precisely the first person discourse via the results of

his experiences and which he writes in a notebook, one that he carries (and therefore teleports) with him always. This personal account will be the same in both the simple teleportation and teleportation with delay cases, something like: "OK, that worked, the GPS confirms that I am in Marseille, I will now leave the booth". By contrast, for an external observer (external to the teleportation booths), the teleportation experiences with and without delay will be very different. With delay, the experience will last a year from the perspective of the third person.

Lemma 1. Observe (this is our first invariance result) the reconstitution delays are not 1-observable, i.e. not first person observable.

We know, with the assumption of 3-locality, that if a candidate survives a teleportation experience, from Paris to Washington, to take another example, he survives independently of all computational activity or even in fact of every event sufficiently distant from the reconstitution.

Consider therefore the following even more delicate experiment. The candidate, after having been properly scanned in Paris is reconstituted in Washington and simultaneously in Moscow. The information, with computationalism, is purely numerical, and can therefore be perfectly duplicated, just like the amoeba. The 3-locality assumption entails that the candidate survives. But where? It is here that it is important to distinguish between the third person discourse and the possible first person discourses. From the perspective of an outside observer (third person), the candidate survived in Washington *and* Moscow. The candidate himself, in the case where he was forewarned of the double reconstitution, could say to himself that he will be, after the experience, in Washington *and* Moscow. But in this case, he is speaking of himself in the third person. In his personal notebook, which is itself duplicated so that he takes it with him, he must note the result of self-localisation of the GPS system in the booth where he was reconstituted. After the duplication, each reconstituted person obtains a unique and precise result: either Washington, or Moscow. A person's notebook contains the mention "I was reconstituted in Washington", and the other "I was reconstituted in Moscow". Neither will contain "I was reconstituted in Washington and Moscow". The one who ended up in Washington could believe *intellectually* that he has a doppelgänger reconstituted in Moscow, and vice versa. This knowledge though is intellectual, communicable in the third person only, and not directly accessible as subjective, private knowledge in the first person.

Concerning such a double reconstitution, if we ask a candidate: "Where will you subjectively feel yourself to be after the experiment", he must recognise that he cannot——as always with all of our hypotheses——experience

surviving in two places simultaneously. As he admits to surviving teleportation and therefore (by 3-locality) to duplication, he must recognise that he is going to experience surviving in Washington *or* Moscow. He must recognise that he is going to write in his notebook "Washington" or he is going to write "Moscow"; in no circumstance will he write "Washington *and* Moscow".

So, he must recognise, unless one reconstitution is arbitrarily privileged, that he cannot predict with certainty the duplication result that he will experience.[4]

From the third person perspective, the situation is perfectly in agreement with the 3-determinism[5] usually associated with Mechanism. But this is the 3-determinism that makes the two reconstitutions numerically identical, and also entails a strict indeterminism from the reconstituted person's point of view. In summary, computationalist 3-determinism entails a *first person* indeterminism, naturally called 1-indeterminism.

Lemma 2. 3-determinism entails 1-indeterminism.

Another remarkable fact is the existence of a form of non-locality. In the same way that 3-determinism entails 1-indeterminism, 3-locality entails a form of 1-non-locality. In effect, distant events cannot change the truth of survival after reconstitution, but the distant event of an identical reconstitution could, from the first person's perspective, change his hope of surviving in such-and-such a place. So distant events could change local predictions in the first person perspective. For example, if a distant cosmic phenomenon were to reconstitute you in a physical state *computationally* equivalent to your actual state, you would have to take this into account if you were to predict your next subjective experience. We could say that an absence of annihilation is equivalent to an annihilation survived by immediate reconstitution (with zero delay). This, I will make more precise below.

We have:

Lemma 3. 3-locality entails 1-non-locality.

Let us consider for the moment an experiment that mixes 1-invariance under delay and 1-indeterminism. The candidate is briefed anew before entering the scanning-annihilation booth where he will be reconstituted in Moscow and Washington. A consequence of lemma 1, 1-invariance under delay, is that from the subject's point of view, these two experiences aren't distinguishable. The upshot is, no matter what method of quantifying the in-

[4] See the thesis for more details, see also my article "Informatique théorique et philosophie de l'esprit"[39].

[5] Henceforth, I will use expression of the form 1-*something* or 3-*something* to designate the something considered from the first or third person respective perspectives.

determinism involved in an experience of self-multiplication is used (whether a probability distribution, or a mass distribution of belief, or something else), this must be invariant under the introduction of delays. For example, if we quantify the domain {Moscow, Washington}, seen as the set of possible future experiences of his consciousness, with a uniform probability distribution in the case of duplication without delay, *therefore* we must use the same uniform probability distribution over the domain {Moscow, Washington} in the case of duplication *with* delay. The delays between reconstitutions do not change in the slightest way the 1-experiences.

We can apply this principle to the thought experiment of teleportation *without* destruction of the original. In this experiment, you are teleported from Brussels to Lille (for example), and as you are not destroyed in Brussels, a third party will see you in Brussels and Lille. Teleportation without destruction of the original is equivalent to a duplication. In particular, it is a duplication where one part has zero delay (Brussels). Not being destroyed is equivalent (under computationalism) to being destroyed then reconstituted *without* delay. Therefore, if you quantify {Moscow, Washington} in a certain way in the experiment of duplication with destruction of the original, we must quantify {Brussels, Lille} in the same way, in the simple teleportation thought experiment without destruction of the original. This will be used later, labelled in the form of lemma 3b. It is lemma 3 with an *absence of destruction* explicitly interpreted as destruction survived by reconstitution without delay.

Real/virtual 1-invariance.

A final invariance principle is required to conclude with the Universal Dovetailer. This last point is hardly truly original: it is connected to the old metaphysical argument of the dream espoused by Hindu logicians; Chinese Taoists; Plato, notably in the Theaetetus; Descartes; Berkeley; Borges; Lewis Carroll; etc. Roger Caillois wrote a nice little book on the argument "L'incertitude qui vient des rêves" (The uncertainty that lives in dreams). According to this argument[6] whenever you're awake, you cannot be sure that you are awake. With the Computationalist Hypothesis, we can substitute, for this argument, dreams by virtual reality. If we simulate a neighbourhood of an environment to a sufficiently high precision on a computer, a first person experience cannot distinguish between the real neighbourhood and this virtual neighbourhood. This sufficient level of precision exists thanks to the existence of a numerical level of substitution granted by the hypothesis. The replacement of real neighbourhoods by virtual ones does not change the

[6] I examine this argument is detail in my 1995 IRIDIA technical report[41].

1-experiences.

All these invariance principles taken together gives the general invariance lemma:

> **General Invariance Lemma:** The means of quantifying 1-indeterminism in the self-multiplication experiments is independent of the 3-places and 3-moments of the reconstitutions, as well as the real or virtual nature of the reconstructions.

The Reversal. I will introduce a new supplementary hypothesis which will replace all the others. I will always suppose that there is a concrete universe but I will also suppose that there is a concrete universal dovetailer (CUD), on which it is concretely and fully executed. As such an execution is infinite, this requires that the concrete universe is infinitely extendible in space and time, permitting the CUD to continue running. I will show that the reversal is a direct consequence of Computationalism *accompanied* by the hypothesis of a CUD.

Recall the "real" experience of dropping a pencil onto the desk. I am ready to let go of the pencil. I would like to predict what will happen. In reality, in everyday life, we are going to use an intuitive "theory" of the sort "every time that I let go of an object, it falls, therefore I expect this will be the same: the pencil will fall on the desk." A more sophisticated theory is "my pencil obeys the laws of physics, and in this case a law that all bodies attract...". Here the theory is more precise, and permits us to conveniently measure the initial position of the pencil to ultimately describe precisely the pencil's trajectory. With a concretely and integrally executed UD, a computationalist must recognise that these previous theories are in the end, mysterious. In effect, at the moment when we are ready to let go of the pencil, as seen from the first person point of view, we will effect an experience of self-multiplication without destruction of the original. In effect, the concrete UD will reproduce it an infinite number of times in its state, described at the right level, where the experience of being ready to let the pencil go occurs. In effect, whatever level of description of the state is necessary for reconstruction of survival, the UD will arrive at this state sooner or later (often later in fact) and generate *all* possible computational outcomes. By virtue of the general invariance lemma, to predict our first person future, we must take into account first person indeterminism over the set of all possible virtual states (emulated by a universal machine in the instance of the CUD), independently of the time and place and virtual character of the reproductions.

A priori, this super-indeterminism implied by the CUD is too strong. There are clearly computational histories, machine dreams as I have sometimes called them, in which the pencil, instead of falling, rises up and turns into a flying pig, or a white rabbit with *watch and vest*[7], from the quantification of the indeterminism of the execution of the CUD. There remains the task of extracting the "correct" physics (with the Computationalist Hypothesis) from the possible computational histories/memories generated by the CUD. "Physics" is reduced to a sum (or an integral) over all possible calculations, it is reduced to the search for a measure over possible computational histories/memories.

Remark. We could believe that the preceding reasoning just leads to a form of solipsism (the doctrine that 'I' am the only dreamer). Yet, we can convince ourselves, from duplicating populations of machines, that the fact of first person indeterminism being communicable in the third person within each replicating population. This allows us to introduce a third person indeterminism, which in effect is only a form of first person *plural* indeterminism. We can predict that if a population of machines share a sufficiently deep[8] computational history and if they observe their most likely universal environment at a lower level than the level where the population is multiplied, they will be confronted with a third person plural indeterminism, or if you prefer, "parallel universes", possible computations, or even counterfactuals. The fact that quantum indeterminism seems communicable and verifiable in the third person, *and* the fact that quantum indeterminism *could* be a particular case of computationalist indeterminism makes solipsism even less plausible. Paradoxically, the quantum confirmation of computationalism makes *our* physical reality, the first person plural one, even more solid.

Everything happens just like that described in the science fiction novel "Simulacron 3" by Daniel Galouye, where the hero end up discovering the virtual nature of their environment by closely scrutinising it. Here, the "quantum peculiarities", of which a quantum computer is one of the most illustrative examples, starts to qualitatively indicate computationalism. I return to the thesis for more commentary on this subject.

Exercise: show that COMP + CUD entails a form of immortality. Discuss.

There remains the elimination of the CUD to finish the demonstration. Suppose that we manage to extract a unique measure over the relative computational histories/memories allowing a precise quantification of the in-

[7] See "Alice in Wonderland"

[8] By deep I mean essentially "the issue of a sufficiently long calculation". We could make this more precise with Bennett's notion of logical depth[6]. Consult the 1995 IRIDIA Technical Report[41].

determinism (and determinism), which permits us to recover the laws of physics. In this case, because these laws belong in such a way to the necessary discourse of Universal Machines (self-referentially correct, honest) relative to their most likely computational history, and an elementary application of Occam's razor and Arithmetical Realism allows us to economise on the CUD and CU hypotheses: we have no need to postulate the existence of a concrete universe, nor of a concretely executing dovetailer at it's core to justify the beliefs and likely observations of Universal Machines. The success of this solution rests on the success of extracting qualitative and quantitative laws of physics from the relative computational histories/memories. *Personally*, I think that the appearance of first person indeterminism (notably the plural one), of forms of non-locality, but especially the appearance of a quantum logic from the Gödellian logic of possible universal machine discourses, explained briefly in the following chapter, not to mention other facts mentioned in the theses, are encouraging facts in this regard. Occam's razor is sufficient here.

From the strictly deductive point of view, we can do without Occam's razor and the "empirical" confirmation by Quantum Mechanics, to eliminate the CUD. We must use the movie graph argument or Maudlin's argument.[42, 41] Independently, Maudlin and I showed the incompatibility of Materialism and Computationalism. As Maudlin postulated materialism, he refuted computationalism. As I postulated computationalism, I refuted materialism [39, 43]. I refer you to chapter 4 of the thesis[42], or the 1995 IRIDIA technical report[41], or Maudlin's 1989 article[43]. To speak frankly, I am no longer satisfied with my presentation of the movie graph argument. In certain respects, Maudlin's presentation is better and more informative. Maudlin seems to ignore Church's thesis, and seems to ignore the *a priori* non-triviality of machine discourses on their possible histories. He did not notice that his argument does not depend on the level of substitution. That is why I say he missed the reversal.

Physics has been reduced to a search for a measure over the possible computational histories/memories. The demonstration was construction from a minimum of "popular" psychology, without which no page of this book would make much sense. Thanks to this popular psychology, the reduction of physics to psychology does not require us to define exactly what we mean by *histories/memories*.

At present, there is a difference between showing that physics must be reduced to psychology, and showing *how* to reduce physics to psychology. I reiterate that psychology is (re)defined by the self-referentially correct discourse of machines. The following chapter suggests a way to proceed. It

isolates an "exact" psychology, that methodically eliminates popular psychology, and allows us in the end to extract a general structure of physical propositions. It is *necessarily* a little more technical. After a fashion, we are going to "interview" the Universal Machine.

Chapter 8

The Machine and its Guardian Angel ("1971" reprise)

Men are naturally moved by two sorts of arguments,
those on one hand that are demonstrative, and on the other,
non-demonstrative.
Averroès, Middle Commentary on the Poetics[13] [1]

One might wonder whether I haven't been inconsistent. Am I not in the middle of relating the amoeba's secret, in the form of Computationalism (I survive a blow to the head, I survive teleportation)?

The intuitive solution, that of Ames and Wyler, is to undergo the secret question, or challenge. The experience of thinking about self-duplication without annihilation of the original should convince oneself that such a "scientific" experience cannot prove—ie communicate in the third person the hypothesis of Computationalism, particularly in the form of a constructive proof of the existence of an adequate level of substitution.

This gives a curious status to the premise: it is necessarily questionable or is an absolute premise, and even more, a necessarily hypothetical hypothesis. Formally, you could be justifiably afraid of building on such shifting ground.

[1] Translation here from French original *Le secret de l'amibe*, citing Ali Benmakhlouf in his book *Averroès*, Les Belles Lettres, Paris, 2000. citing Butterworth, C., Harîdiî A. A. (eds), Le Caire 1987.

A similar critique is often addressed to those who talk about the incommunicable, or speak of the ineffable.

As the young (and very positivist) Wittgenstein espoused in his "Tractatus" the well-known aphorism "Whereof one cannot speak, thereof one must be silent", one is right to wonder of what he can speak. And can he speak, if he must keep silent?

Even Lao Tzu never missed the opportunity to remain silent when he affirmed that the Tao with a name is not the Tao?

The base of the simplest and most naïve idea, which I glimpsed in 1971, is to interview the Universal Machine, modelling or capturing, in effect, honest communication by formal proof. At this time, certainly—see the preceding chapters—I didn't know of Church's thesis, and I hadn't grasped the consequences of Gödel's theorem for all machines; I thought often of an entity like Russell and Whitehead's PRINCIPIA MATHEMATICA, or formal Peano arithmetic (PA), as being the veritable *Escherichia coli* of research on self-reference.[2]

I am not going to interview any old universal machine. I am only going to interview those that are self-referentially correct, and in particular, consistent.[3]

The fact that these machines are universal is, to a certain point of view, equivalent to the fact that they know how to prove all Σ_1 (pronounced "sigma one") expressions, provided they are true, of course.

A proposition is said to be Σ_1 if it is (provably) equivalent to a proposition of the form $\exists n : P(n)$ with $P(n)$ being "mechanical", communicable, verifiable or falsifiable.

In effect, what I was going to discuss here goes for all formal theories, or machines, capable of proving a sufficient number of the theorems of elementary arithmetic. Gödel discovered for these machines, it is always possible to

[2]This chapter supposes a minimum of knowledge in classical propositional logic. See, for example, the remarkable small book "Introduction à la logique"[52]

As for modal logic, one could consult the book by Jean-Louis Gardiè[27]. A classical treatus is the book by Chellas[14]. Obviously, one could also consult my thesis[42], or the chapter "Théologie et Modalité" in the 1995 IRIDIA technical report[41].

There is also in English a *recreational* introduction, to Gödelian modal logic, of self-reference and provability you could also say, by Raymond Smullyan [57]. Raymond Smullyan is the author of a great number of technical, recreational and philosophical books that all illustrate the profoundness of Gödel's incompleteness results. For the logic of self-reference, or provability, the classics are by Boolos[8, 9], as well as Smoryński[56]. Rucker[54] is another captivating introduction to the incompleteness theorem.

[3]A machine or formal theory is consistent if and only if it cannot prove false propositions.

translate the proposition "p is provable by me" or "I know how to prove p" into the language of the machine. One can identify this language to a portion of elementary arithmetic through an appropriate intermediate coding.[4]

"provable $\ulcorner p \urcorner$", which I will denote $\Box p$, can perhaps be defined by an arithmetical formula, even Σ_1: "there exists an n such that n is a numerical representation of a proof of $\ulcorner p \urcorner$", where $\ulcorner p \urcorner$ designates a numerical representation of proposition p. $\ulcorner p \urcorner$ is called the Gödel number of proposition p. Effectively, it is mechanically verifiable whether or not a number n is the Gödel number of the proof of proposition p (which has Gödel number $\ulcorner p \urcorner$).

These machines are automatically subject to Gödel's diagonalisation lemma,[5] as I have spoken of in chapter 3. In particular, there are propositions p such that that machine can prove $p \leftrightarrow \neg\Box p$. It is easy to convince oneself that p is automatically true and *non* provable for the consistent machine. In effect, if p, which is equivalent to $\neg\Box p$, were false, $\neg p$ would be true, but $\neg p$ is equivalent to $\Box p$, and so p would be false and provable, and the machine inconsistent. We can see, therefore, that there are true non-provable propositions for consistent Universal Machines capable of proving sufficient elementary arithmetical theorems. This is Gödel's first Incompleteness theorem from 1931. The second Incompleteness theorem, which I will return to later, affirms that the consistency of the Machine $\neg\Box\bot$ is one such proposition, true, yet not demonstrable by it.

If we bet that we are such Machines, we have therefore a means of communicating about that of which we can or cannot speak; exactly what we are researching. I identify the honest or scientific communicability of the machine with the formal provability predicate of that machine. In this sense, as John Myhill says, Gödel's Incompleteness theorems are the first theorems of an exact psychology. A consistent machine cannot give a formal proof of its own consistency.[6]

As well, we are going to be particularly interested in machines that possess more than a minimum of introspective capability. They not only know how to prove all true Σ_1 propositions, they also know how to prove

[4]We suppose, of course, that a level of description has been fixed for the machine, for example, corresponding to the level of survival under substitution.

[5]This is true even without their introspective quality described above.

[6]Note that the result is more general and concerns also machines having access to oracles (in the sense of Turing 1939), that is to say, infinite data which they may consult if necessary. Most likely, this chapter will exceed the limits of the hypothesis of Computationalism. Strictly speaking, Computationalism appears in this chapter when the interpretation is restricted to propositional variables from Σ_1.

the sense of what they know how to prove, those Σ_1 p:

$$p \rightarrow \Box p,$$

where $\Box p$ represents the internal predicate of provability. We say that they know how to prove their own Σ_1-completeness. They know or can know that they are (at least) universal.[7]

We have seen that the predicate "provable($\ulcorner p \urcorner$)", abbreviated by $\Box p$ could be translated into the language of the machine by a Σ_1 proposition. Therefore, for these machines, we know, thanks to their introspective quality, that they know how to communicate whatever the proposition p is in the language of the machine (or in the language of arithmetic):

$$\Box p \rightarrow \Box\Box p$$

Note that, since $\Box p$ represents the *arithmetical* proposition "provable($\ulcorner p \urcorner$), $\Box\Box p$ represents the arithmetical proposition "provable($\ulcorner$provable($\ulcorner p \urcorner$)$\urcorner$)".

The logician Léon Henkin posed a very interesting and completely natural question: what about the self-referential propositions p that, instead of affirming their own non-provability like Gödel's sentence, by contrast affirm the own provability? Such propositions exist by virtue of the diagonalisation lemma. *A priori*, these propositions could be false and non-provable, or true and provable. No contradiction appeared, differing from the Gödellian proposition which affirms its own non-provability.

In 1955, the Dutch logician M. H. Löb published an article with the solution to Henkin's problem. Just as peculiar as that Henkin's propositions could appear to affirm their own provability, is the fact that they are in fact *always* true and provable.

Gödel's theorem (the existence of true and non-provable propositions) is built on a version of Epimenides's paradox. Gödel replaced Epimenides's proposition "I am false" with "I am not provable" expressed in the language of the machine. Even Löb's proof uses a curious little self-referential paradox. Here, in effect, is a proof of the existence of Santa Claus! (Find the error!) It consists of a small amount of propositional logical reasoning. Note the double utilisation of the rule of *modus ponens* MP : $\frac{A\, A \rightarrow B}{B}$, which says that if A has been demonstrated, and you can demonstrate $A \rightarrow B$, then you can deduce B.

Consider the following sentence or the proposition P:

[7] Ultimately, "may be able" will be taken in the weak sense of "true and communicable", and it must be said only that for each Σ_1 proposition p, the machine knows how to communicate $p \rightarrow \Box p$.

If this sentence is true, then Santa Claus exists

"This sentence" designates the whole sentence P, it is therefore self-referential.

I will first prove P. P is a conditional proposition, and the premise of P is P itself. To prove a conditional, you assume the premise, and show that the rest follows. Suppose that the premise is true, that is suppose that P (this sentence) were true. Then P is true, and automatically "P entails the existence of Santa Claus" is true. So therefore, with the hypothesis P, Santa Claus exists by MP. We have shown that if P is true, then Santa Claus exists. But that is exactly what P states. Therefore, we have shown, without supplementary assumptions, that P is true.

At present, P just states that if P is true, then Santa Claus exists. Now, we come to showing P. By a new application of MP, we have that Santa Claus exists.

Where is the error?

Many think that the error resides in the use of a self-referential proposition. But we know that sufficiently rich universal machines obey the diagonalisation lemmas, and that self-referentiality is inescapable. What is going on?

Recall the Tarski theorem from chapter 3: One cannot translate the predicate of truth *of* the machine *into* the language *of* the machine. Löb's sentence, P, is simply not translatable into the language of the machine. This resolves the paradox, at least with respect to the world of machines.

Gödel proved his Incompleteness theorem by replacing truth with provability. Similarly, Löb proved[8] his theorem, which resolves Henkin's question, by replacing truth with provability in the sentence P. Löb showed, in effect, that if a machine proves $\Box p \rightarrow p$, it proves p. That gives the response to Léon Henkin's question, as the statement $p \leftrightarrow \Box p$ entails $\Box p \rightarrow p$. So if the machine proves the former, it proves the latter, and we can apply Löb's theorem. This is truly astonishing: this resembles a form of "wishful thinking"[9] or the Coué method: if I prove that if I had proved p, I would have p, therefore I have proved p. Strange, but true. And not only is that true, but we can show that machines can also prove this result. In fact, by a proof that mirrors the reasoning of the Santa Claus paradox, machines can show:

$$\Box(\Box p \rightarrow p) \rightarrow \Box p.$$

[8]This proof, in contrast to Gödel's, requires the supplementary capacity of introspection described above.

[9]Taking one's desires for reality.

This is Löb's formula. It is often interpreted as a manifestation of machine modesty. It communicates a proof of p entails p only when it proves p.

Now, $\neg p$ is equivalent to $p \to \bot$ where $\bot$ stands for the generic false proposition (you could replace $\bot$ by $0 = 1$ everywhere; and $\top$, which stands for the generic true proposition, could be replaced by $1 = 1$). In replacing p by $\bot$ in Löb's formula, you find Gödel's second incompleteness theorem as a particular case:

$$\Box\neg\Box\bot \to \Box\bot,$$

which you can also write

$$\Diamond\top \to \neg\Box\Diamond\top,$$

where $\Diamond p$ is the usual abbreviation of $\neg\Box\neg p$. Even $\Box p$ is equivalent to $\neg\Diamond\neg p$. In this case, "I am consistent", that is "I cannot prove false", or $\neg\Box\bot$, becomes $\Diamond\top$. Gödel's second theorem is provable by the machine itself, and means "if I am a consistent machine, I cannot demonstrate that I am a consistent machine". That shows also that inconsistency $\Box\bot$ is also consistent! The machine that does not prove false can prove false. It could be wrong or dreaming. We have, in respect of the machine, $\Diamond\Box\bot$.

There are true propositions of the machine that the machine can prove, and other true propositions, that always concern the machine itself, that the machine cannot prove. With Gödel's second theorem, we see that the machine can hypothetically justify that it can't prove certain propositions, including the important proposition of self-consistency, $\Diamond\top$ that we can interpret liberally, largely and with a grain of salt by "I cannot communicate the false statement", "I am awake", "I am honest", "I am intelligent" or "I am conscious".[10]

Let us call "modal propositions", the propositions of elementary propositional logic extended with the connectors $\Box$ and $\Diamond$. When the variables are

[10]Dostoyevsky would gave defined consciousness as "Consciousness is the presentation of accessible truth for a man" ("La conscience, c'est le pressentiment de la verité acessible par un homme"), cited by Oleg Tabakov, man of the Soviet theatre, himself cited by J. P. Thibaudat[61]. I haven't found the original text. In using Kripke's geometric interpretation of modal formulae (see the appendix on modal logic in my thesis), we can express this definition (axiomatic and partial) in a more everyday fashion. Note well that consciousness, honesty, etc. are not identified with consistency. It is just suggested that there are modal axioms capable of capturing common aspects of these notions. With an even larger interpretation of $\Diamond p$ as "I am alive (surviving)", the modal formula corresponds to Gödel's second theorem, expressed for being alive, ie able to die. Everything here is has been considerably developed in the 1995 IRIDIA technical report[41].

substituted by propositions in the machine's language, and when $\Box p$ is interpreted, as one has done just here, by propositions of the style provable($\ulcorner p \urcorner$), carried out in the machine's language, we could ask if there is a formal modal logic theory capable of axiomatising correctly and completely an interview with the machine.

This question, asked by George Boolos, will be resolved in the affirmative by Solovay in 1976[58]. Solovay showed, in effect, that the modal logic system, which he called G, described below, axiomatises the integrality[11] of the self-referentially correct machine's discourse (or formal provability in sufficiently rich theories). G is given by the following axioms and rules:

AXIOMS:	$\Box(A \to B) \to (\Box A \to \Box B)$	K
	$\Box A \to \Box\Box A$	4
	$\Box(\Box A \to A) \to \Box A$	L
RULES:	$\frac{A \quad A \to B}{B}$	MP
	$\frac{A}{\Box A}$	NEC

NEC denotes the inference rule called necessitation: if I have shown p, then I know how to show $\Box p$. K comes from Kripke, "4" is the accepted name (but rather silly) of the formula $\Box A \to \Box\Box A$. L comes from Löb, certainly.

But what is "incommunicable, but true"? We know that there are true propositions by the machine that the machine doesn't know how to prove, like self-consistency $\Diamond\top$, consistency of inconsistency $\Diamond\Box\bot$, "self-correctness" $\Box p \to p$, etc.

Solovay offers an unexpected present: the set of true modal propositional formulae, provable *or not* (always interpreted in the machine's language) is *also* completely axiomatisable. In particular, the following system G*, captures the set of true modal propositions, communicable and incommunicable, concerning the machine:

AXIOMS:	all the axioms of G,	
	$\Box A \to A$	T
RULES:	$\frac{A \quad A \to B}{B}$	MP

Note the loss of the necessitation rule. It is an easy exercise to show that G* + the necessitation rule gives an inconsistent system.

[11] Only at the propositional level. Russian logicians have shown the *high* undecidability of the logic of first order self-reference. These proofs are detailed in George Boolos's 1993 book[9].

Solovay showed the adequacy and completeness of G and G* for proof about and by the machine, and the truth about the machine respectively, but he also showed the decidability of these systems: G and G*, like the propositional logic that can be generated[12] "mechanically". G* extends G. The corona G*\G, becomes a decidable system, closed for *modus ponens* (see the thesis), capturing the incommunicable propositional truths of the machine, the infinite spaces of the amoeba's secret.

G axiomatises faithfully and completely the discourse of the consistent machine (honest and/or awake). Therefore, sometimes I identify G with the discourse of *the machine itself*, it being sufficient to recall the way in which the symbols are interpreted in the machine's language. In the same way, and in honour of Judson Webb (I explain why in chapter 6), I call G* the discourse of *the guardian angel of the machine*. G* doesn't speak of itself, but speaks of G, or about the machine. G* axiomatises the part, as well as the communicable (G* extends G) but also the incommunicable truths *of* the machine. We can therefore interview the Universal Machine, and also its guardian angel.

The third person. If you model, or even if you capture honest communication of consistent machines by formal provability, you never leave third person scientific discourse. Of course, there is a self-referential discourse, and when the machine communicates $\Box p$, it is well on the way to correctly proving that it can prove p, but this self-reference is a third person self-reference. When the machine communicates $\Box p$, it communicates an arithmetical proposition (for example), or a proposition in its machine language that is equivalent to "there is a number (a list, a sequence of signs) which codes the demonstration of a proposition coded by $\ulcorner p \urcorner$". *We* know, eventually, if the machine is not too complicated, that it is correct and indeed, self-referentially correct. But Gödelian self-reference is quasi-accidental according to the machine's natural behaviour. Clearly, this code, extending the description to a formal level, values everything equally as the duplicate of the machine, as its eventual doppelgänger after an experience of self-duplication.

Therefore act. And, with a bit of luck, that gives, *by construction* an honest (scientific) discourse, guaranteeing that the entirety of our conversation, as *philosophical* as it may seem, admits an interpretation in terms of true arithmetical propositions *of the machine* and provable *by the machine*,

[12]See my technical report from IRIDIA[41] for demonstrations of theorems (in LISP) for G and G* and other logical systems in this chapter, and many other considerations of the logic of self-reference.

and eventually true *about the machine* but *not* provable *by the machine* when interviewing the guardian angel. This is also speaking in the third person, its discourse is also scientific, but only on questions of the machine itself.

If the guardian angel can communicate non-communicable propositions, it is because these are non-communicable by the machine in reference to that of which it can speak. G*, by contrast to G, does not speak of itself, instead it speaks about the machine. The guardian angel can say $\Diamond\top$, which doesn't mean "I am consistent", but that *the machine* (*the machine which it guards*, if you like) is consistent. This is the mismatch between the machine and its guardian angel that clarifies in a brilliant way, in the world of machines, the possible discourses of the incommunicable.

All of these remarks extend naturally to the propositions of psychology or of physics, this terminology giving rise to a new interpretation imposed by The Reversal. All that is needed is to translate the terminology of psychology and the physics of machines into the terminology of formal communicability by the machine or its guardian angel.

Put another way, to capture the logic of the first person, we should, in the interview with the Universal Machine and its guardian angel, translate the "I" of the subject that knows, measures or observes, in terms of the third person provability formula. I will get to this shortly.

The first person knower. The first person is all about the one who knows. It is the subject of knowledge. Modal formulae typically used to axiomatise knowledge have the reflection form $\Box p \to p$ in logics closed for the necessitation rule $\frac{p}{\Box p}$ (in particular we want $\Box(\Box p \to p)$. It guarantees in a certain way the umbilical reattachment of the subject to truth, and at the base. Note that we restrict ourselves again to knowledge concerning communicable propositions. We require that knowledge of p entails the communicability of p.

Gödel, in his little 1933 paper[28], had already noted the inadequacy of the formal provability predicate for capturing knowledge,[13] because $\Box\bot \to \bot$ is incommunicable, though true, and certainly for this reason, $\Box(\Box\bot \to \bot)$ is false. In effect, the guardian angel warns us, just as surely as it affirms the machine's honesty ($\Diamond\top$), that the machine can (ie is consistent with) also communicate false ($\Diamond\Box\bot$). We could therefore also interpret provability as a *belief*, because falsity is possible, in contrast to knowing. For example, "Dominique *believes* the Earth is flat"; you never say "Dominique *knows* the Earth is flat". Knowledge is therefore connected to truth by definition. In identifying formal, communicable proof with provability $\Box$ of the Universal

[13] This had already been developed by Kolmogorov[33]

Machine, we put scientific propositions and third person discourses clearly in the camp of the believed ... and *accidentally* known. The guardian angel knows that the modesty of scientific discourse of the machine is logically attached to the possibility of error, lying or dreaming (usual, non-lucid). The machine does not believe it, but it can infer it, and raise doubts.

If the formal provability formula does not verify natural axioms of knowledge, we must try to define knowledge from the provability formula. The simplest way of attaching provability to the umbilical cord of truth would be to define "p is knowable" by "p is provable *and* p is true". In replacing "provable" by "opinion" or "justifiable opinion", we recover the attempts at definitions of knowledge that Theaetetus proposed to Socrates in Plato's Theaetetus.[14] But this definition isn't expressible in the language of the machine. This is a new consequence of Tarski's theorem, we do not know how to define "p is true" *for* the machine, *in* the language of the machine itself.

That seems impossible. In a very general way, it is easy to show that it is impossible to define, for a consistent machine, an arithmetisable predicate [15] satisfying both the reflection formula, $\Box p \rightarrow p$ and necessitation $\frac{A}{\Box A}$. In effect—by direct consequence of the diagonalisation lemma—the machine, for a certain proposition k, would prove $k \leftrightarrow \neg\Box k$. In particular, it proves $\Box k \rightarrow \neg k$. Since it proves reflection $\Box p \rightarrow p$ for all propositions p, it therefore proves in particular $\Box k \rightarrow k$. By propositional calculus, it therefore proves $\Box k \rightarrow (k\&\neg k)$, ie $\Box l \rightarrow \bot$, which is also $\neg\Box k$. As it has already proven $k \leftrightarrow \neg\Box k$, it proves k, and by necessitation, proves $\Box k$. Therefore it proves $\neg\Box k$ and $\Box k$, and so is inconsistent.

For our machines with sufficient introspective ability, and which verify Löb's theorem, and even prove it, we see even more quickly that reflection combined with necessitation is forbidden, because the application of necessitation on reflection (with p substituted by $\bot$) gives $\Box(\Box\bot \rightarrow \bot)$, which by Löb and MP gives $\Box\bot$, which by a new application of reflection gives $\bot$.

Just as Tarski's proof shows that some truth of the machine isn't expressible or definable in the language of the machine, the little reasoning above shows that knowledge, and in a very general sense no matter how, axiomatised by reflection and closed under necessitation, is no longer expressible (arithmetisable) in the language of the machine.

Note that G and G* are coherent in this regard: G is closed under

[14] And in many other discussions, currently, see for example, Burnyeat 1991 in the nice collection of Monique Canto-Sperber[12]

[15] or definable in the language of the machine, and therefore subject to the diagonalisation lemma (see chapter 3)

necessitation, but cannot prove reflection, and G* proves reflection, but doesn't verify necessitation (forcibly so, because if it obeyed necessitation, that would entail that G, the machine itself, could prove everything that the guardian angel can prove. The corona would be empty, and the *amoeba* or the universal machine, wouldn't have any secrets. But this also shows that neither G, nor G* captures *directly* a description of the "knower" or the first person).

A variation of the argument presented here is often presented as a means of using Gödel's theorem to distinguish man and machine. In simple terms, knowledge is not arithmetisable, ie translatable into the language of a machine, therefore machines don't have a knowledge predicate, and cannot know.[16]

In reality, the argument shows only that knowledge cannot be arithmetised, whether by machine, or anything, or anyone.

But how can we go about interviewing the Machine or its guardian angel concerning knowledge if we cannot translate the concept of knowledge into the language *of* the machine?

Well, a very simple way[17] is the following. In place of defining knowability of p by "p is provable and p is true". Always taking inspiration from Theaetetus, we will define knowability of p by "p is provable and p". This allows us to avoid the impossible usage, as we have just seen, of the truth/knowledge predicate. We are going to simply define a new modal connector at the level of propositional logic, $\boxdot$, where $\boxdot p$ is directly interpreted as $\Box p$ & p. We replace the impossible usage—due to Tarski—of TRUE($\ulcorner p \urcorner$) by the simple assertion p.

Note that reflection of this knowability is not only true of the machine, but is provable *by* the machine: the machine proves $\boxdot p \rightarrow p$, because it obviously proves $(\Box p \& p) \rightarrow p$.

Even the logic of the machine's discourse on $\boxdot p$ is closed under necessitation. In effect, with necessitation $\frac{p}{\Box p}$ and the rule $\frac{p}{p}$ (that is itself derived from $p \rightarrow p$ with modus ponens), it pertains that if the machine proves p, it proves $p \& \Box p$ and so is closed under $\frac{p}{p \& \Box p}$, that is $\frac{p}{\boxdot p}$.

There is no paradox: $\boxdot$ is *not* arithmetisable. Even though Tarski's

[16] This is also connected to the paradox of Kaplan and Montegure 1961, and to Plato's knowledge paradox (cf Monique Canto-Sperber. See [12]).

[17] Discovered and studied independently by many logicians, the American, George Boolos 1980; the New Zealander, Robert Goldblatt 1978; and the Russians Kuznetsov and Muravitsky 1977, however, in a more extended context of the paradoxes of knowledge. Artemov 1990 proposed as a thesis the equivalence of intuitive or informal provability with $\Box p \& p$. See the 1995 IRIDIA technical report[41].

theorem shows that there is no truth predicate definable in the language of the machine, and that just means there is no $V(x)$ that the machine might prove $V(\ulcorner p \urcorner) \leftrightarrow p$. That there isn't a Theaetetical knowledge predicate definable in the language of the machine simply means that there isn't a $C(x)$ so that the machine proves $C(\ulcorner p \urcorner) \leftrightarrow (p \& \text{provable}\ulcorner p \urcorner)$. Knowledge, like arithmetical truth, is not subject to the diagonalisation lemma because it is not definable in pure arithmetical terms. In defining truth of p by the pure assertion p, we discover a non-trivial discourse of the machine on knowability that occurs in some sort of arithmetical definition, and which occurs on the representation (Gödel numbers) of the formulae of which it speaks.

Boolos, Goldblatt, as well as Kuznetsov & Muravitsky at the end of the 1970s, independently showed that this logic of knowability is completely axiomatised in a correct (sound) way, and complete, by a system invented in 1967, by the Polish logician Grzegorczyk. It consists of the standard logic of knowledge, knowability in effect, S4 (K, T, 4 + the MP and Nec rules), to which is added Grzegorczyk's slightly curious formula Grz:

Axioms:	$\Box(A \to B) \to (\Box A \to \Box B)$	K
	$\Box A \to A$	T
	$\Box A \to \Box\Box A$	4
	$\Box(\Box(A \to \Box A) \to A) \to A$	Grz
Rules:	$\frac{A, A \to B}{B}$	MP
	$\frac{A}{\Box A}$	NEC

Geometrically, with Kripke semantics,[18] we can see this logic as a sort of temporal logic describing the future evolution of states of knowledge developing irreversibly (anti-symmetrically). Thanks to a suggestion of Gödel's[19] where S4 can emulate by the intermediary of a modal transformation (see the thesis) intuitionist logic (Brower's subject logic, formalised according to Heyting), Kripke discovered his (well known) semantics of intuitionist logic. That describes the same temporal evolution of states of knowing. This is a logic of "subjective" time.

We can demonstrate that the discourse of the guardian angel on knowability brings up no more than the discourse of the machine itself. S4Grz*, the collection of true modal propositions, and therefore provable by G* (thanks to Solovay's theorem) is equal to S4Grz. From the point of view of knowability (arithmetic), truth is equivalent to provability. See Boolos 1993.

It is the results of these transformations which permitted Goldblatt to

[18] See the modal logic appendix[42].

[19] [28]. This suggestion was proved by McKinsey and Tarski in 1948[44].

extract a purely arithmetical interpretation, hence interpretable in the language of the machine, of Intuitionist Logic IL. If I discuss this result, it is not because it is very interesting, but because I'm going to be inspired by it to question G and G* on the Universal Dovetailer, and the logical origin of beliefs in physical propositions. Note that here also IL=IL*: from the vantage point of the intuitionist subject, truth is equivalent to simple assertion.

We could also interrogate G, the machine itself and G*, the guardian angel, on mixed propositions. In particular, for every proposition p (in the language of the machine), G* proves $\Box p \leftrightarrow \boxdot p$, but the machine does not.[20] This illustrates the distinction between the first person and the third person and is an intensional (modal) nuance of provability. It consists of different points of view of, and on, the same machine.

The first person who measures, or who senses. There are all sorts of remarkable things about these intensional nuances that do not depend on the fact that they are vaccinated against diagonalisation, like $\boxdot$.

In substituting "assertable" truth of a proposition p, by the *possibility of truth*, that in passing from $\Box p \& p$ to $\Box p \& \Diamond p$, we define a new intensional variant that is arithmetisable. It is an arithmetical refinement of Theaetetus's idea.[21]

This leads us to define a new modal connector, $\triangle$, with $\triangle p$ equivalent to $\Box \& \Diamond p$.

That it is an intensional variable is assured by G*. Here also, G* proves $\triangle p \leftrightarrow \Box p$, but G doesn't prove it. G* even proves $\triangle p \leftrightarrow \boxdot p$, but G does not. From the guardian angel's point of view, it is always the same machine (defined in the third person by the propositions it communicates). For the machine's point of view, there are always nuances distinguishing different sorts of viewpoints. The possibility of these nuances has its origin in the incompleteness phenomenon.

"$\triangle$" is clearly arithmetisable, meaning again definable in the language of the machine such that it corresponds to provable($\ulcorner p \urcorner$) & consistent($\ulcorner p \urcorner$).

Therefore the predicate represented by $\triangle$ is automatically diagonalisable, in the subjective sense of the diagonalisation lemma. We can show that the "Gödellian" formula k is provably equivalent to "$\neg \triangle k$ which is

[20] For a use of this fact and facts of this kind for a reflection on dreaming and being awake, see the 1995 IRIDIA technical report[41]. I show there that G and G* assure the coherence of the metaphysical argument of dreams (or of virtual reality) used by Theaetetus, and its use in the context of Computationalism. We will find also a refinement of Slezak's analysis of Descartes's *cogito* argument, as well as variations on Lucas's refutations.

[21] One of the refutations of Lucas by Judson Webb uses a similar idea.

equivalent to $\Box\bot \vee \Diamond\Diamond\top$", where $\vee$ is taken for the usual logical "or".[22]

The interview with G, with the Machine itself, on the subject of modal propositions using the new connector, gives a new modal logic, that I call Z. Even the interview with the guardian angel G* produces a new logic Z*. By contrast to Grzegorczyk's system S4Grz, the modal systems Z and Z* are distinct: $\Box\top$ is provable by Z* but not by Z, because $\triangle\top$ belongs to G*\G. In particular, we see that the logic Z is not closed under the necessitation rule (like G*, but unlike G and S4Grz). This rules out the use of Kripke semantics for trying to axiomatise Z, and in particular the question of even knowing whether Z and Z* are completely axiomatisable remains open. Nevertheless, it is not difficult to use the work of Solovay to show that Z and Z* are decidable and mathematically well-defined.[23]

The refinement of Theaetetus's idea, which corresponds to the passage of $\Box p\&p$ to $\Box p\&\Diamond p$ models the passage from the actuality of p to the possibility of p. This passage is rendered quasi-obligatory by the argument of the Universal Dovetailer, where, I recap, the computationalist who wants to predict her future is obliged to quantify an indeterminism of the first person on the set of *all consistent* extensions. As in the philosophical approach of the indexical to the actual, where actuality is defined by every possibility *seen from the inside*, Computationalism forces the interiorisation of consistency. Furthermore, with a new connector $\diamondsuit$ standing for $\neg\triangle\neg$, we can show that G proves

$$\triangle p \rightarrow \diamondsuit p,$$

which means Z proves $\Box p \rightarrow \Diamond p$. This formula has a standard name amongst modal logicians: D. "D" comes from "deontic". In deontic logics, the interpretation of the modal square $\Box$ is the obligation, and the dual $\Diamond$, that is $\neg\Box\neg$ corresponds to permission (p is permitted if and only if it is not obligatory to have $\neg p$, also that p is obligatory if and only if it is not permitted

[22] With this new form of Gödel's theorem, we automatically obtain the intensional refinements of Descartes's *cogito ergo sum* due to Slezak, and in my technical report, I illustrate how one could use this refinement to criticise the argument of positivist philosopher Norman Malcolm (Oxford) against the existence of conscious experience during dreaming and in machines.

[23] We can also show (see the thesis) that Z admits a neighbourhood semantics (notion attributed to logicians Scott and Montague). In modal logic, we often define the intension of a proposition by the set of possible worlds where this proposition is true. The logical structure of Z's propositions confers a quasi-filter structure on the neighbourhoods. A quasi-filter is a filter without a maximal element. Once we have a filter, we can construct Kripke-type semantics. In this sense, Z features quasi-Kripke semantics that permit the usage of a non-negligible portion of Kripke's intuition. See Challas's book (referenced in the thesis) for Scott and Montague's semantics. See the 1995 IRIDIA technical report[41].

to have $\neg p$). The formula D indicates that either the obligatory must be permitted, or that which is forbidden cannot be rendered obligatory. It is an elementary axiom of rights.

Formula D is also the entry point for research into quantification of any indeterminism. In systems closed under the rule of necessitation, D has been been used to capture the modality of certainty in probability. Without necessitation, it has been used to model or capture notions of credibility. In effect, if p is a certain proposition, we would like $\neg p$ to not be certain. This makes an arithmetical version of Theaetetus's definition, a variant of Gödellian provability touching on a sort of belief anticipating self-consistency. [24]

We can see this in another way. In terms of possible worlds, otherwise known as Kripke semantics, a proposition of the form $\Diamond p$ is true in a world M_1 (or a state or situation, ...) if I can access a world M_2 from M_1 where p is true. Imagine that you end up in a situation without an exit, a world where you cannot access any other possible world, a sort of *cul-de-sac* or dead-end world. In such a world, all propositions of the form $\Diamond\#$ are false, therefore all propositions of the form $\Box\#$ are true. In a dead-end world, nothing is possible and everything is necessary! In attaching consistency to provability, like the arithmetisable version of Theaetetus's idea, we essentially filter out the dead-end worlds. This must be done since the probabilities (or credibilities) that appear in The Reversal are defined on consistent extensions. This justification must be nuanced in light of our arithmetical context (of machines), else the arithmetical version of Theaetetus's idea

[24] With folk psychology, we can admit that "survive" requires remaining conscious. If we regard consciousness as a logical daughter of consistency, by incompleteness it cannot be purely logical. We can therefore see consciousness, in a self-referentially correct machine, as the fruit of an instinctive anticipation (programmed or selected, one level on another) of consistency of itself. The decidability of $G^*\backslash G$ illustrates the inferable character of a non-negligible collection of non-provable propositions. We are not very far from Helmholtz's idea of perception produced by instinctive inference. This generates a sort of voyage from G to G^*.

The machine can infer inductively a proposition of $G^*\backslash G$ and retain it as a secret or bet or simply adorn it with a question mark. But it can also integrate the new proposition by modifying its code (at one level or other). In this case, it changes itself as a teller of truth. G and G^* *automatically* apply to the new machine. In addition, the *arithmetic interpretations* of G and G^* change as well because they apply to a new machine with its new language. This integration suggests a role for consciousness: it is that which permits a relative acceleration of one universal machine relative to another. In effect, inferring and integrating a (relatively) consistent proposition, makes an infinity of undecidable propositions decidable, but also shortens the length of an infinity of provable propositions. I allude here to the speed-up theorem of Gödel. Consciousness of itself develops in a way that the non-communicability of this consciousness/consistency is itself anticipated. That allows the distinction and recognition of self versus other.

loses our Kripke semantics, and notably, the possibility of structuring worlds sporting accessibility-relations. But this justification could be corrected to work with Scott and Montague's semantics. We can see in this idea a sort of Darwinian arithmetic: we interrogate the machine about its exclusively consistent extensions. We can also see there a generalisation of the anthropic principle, a sort of "universal machine-tropic" principle: we interrogate a consistent machine on its possible environments where it remains consistent. We simply forbid quantifying the indeterminism of the "dead-end worlds". In the end, the filtering of the consistent extensions will justify the quantum logic of propositions where a form of consistency is observable ($\Box\Diamond p$) for p accessible by the Universal Dovetailer.

Another and final motivation of a general nature for the passage from S4Grz to Z and Z* follows. We would like to limit the number of extensions which, while consistent, are aberrant; like "hallucinatory" experiences such as flying pigs or white rabbits (with waistcoat and pocket watch).

Like in algebraic geometry, where adding equations to an existing system of equations restrains the set of geometric objects satisfying the system, in formal logic, adding axioms restrains the class of its models.[25] Unfortunately, adding axioms to a sufficiently rich theory and subject to diagonalisation does not truly diminish the number of models due to the infinity of undecidable propositions: we would need to include an infinity of formulae if we wished to definitively rid ourselves of flying pigs in this way. A better idea consists of weakening the logic in the hope of multiplying sufficiently non-aberrant models. In fact, Z weakens S4Grz quite considerably. In this way, we augment the number of models and we can argue that the obtained augmentation with Z (and Z*) produces neighbourhoods having the power of the continuum. All that remains is to isolate a proximity relation on the extensions and to show that (mostly) all our extensions are relatively "normal".

The application of Theaetetus's idea to the logic of self-reference, ie the passage from $\Box p$ to $\Box p\&p$ already defines a space of the knowable that we can see as the psychological reality of the first person.

The arithmetical version of this idea, the passage from $\Box p\&p$ to $\Box\&\Diamond p$,

[25] A model is a mathematical structure which satisfies (renders true) a theory, seen as a set of axioms and rules. Logicians, like painters, use the word model to designate a possible reality. Theory, like the canvas, aims to approximate or capture aspects of this reality. Physicists often use the word "model" for the theory or theoretical approximation, as when we speak of a reduced model or of modelling (for example by the Bohr model of the atom). Perhaps this explains the frequent arguments between logicians and physicists where people talk past each other.

defines a more tangible psychological reality. It is easy to show that neither Z, nor X* proves the formula 4 ($\Box p \rightarrow \Box\Box p$), and we can argue that it makes this reality a sort of immediate belief, not directly accessible to introspection.

Our goal, however, is to isolate physics, or at least the skeleton or logical structure of propositions about "observables". The Universal Dovetailer Argument motivates us to translate certain observations, seen as a sort of immediate belief, by provability accompanied explicitly by consistency. But we haven't always introduced the Universal Dovetailer explicitly in our interview of G and G*.

With Computationalism, *physical* indeterminism isn't defined on all our consistent extensions, only on those that extend the states attained by the Universal Dovetailer. Recall that a universal dovetailer is only a crushed Universal Machine: a catalogue of histories generated and accessible states. Arithmetically, universality can be modelled by Σ_1-completeness, and the Universal Dovetailer could then be considered a catalogue of proofs of (true) Σ_1 propositions. Such propositions are verifiable when true, but not necessarily refutable when false.[26]

We must therefore limit the arithmetical interpretation of propositional variables p to Σ_1 propositions. We know (see above) that the propositions $p \rightarrow \Box p$ are true for these propositions and even provable by a sufficiently introspective machine.[27]

In summary, we obtain phenomenology of matter by effecting:

- the Σ_1 restriction,
- the arithmetical version of Theaetetus's idea.

The result gives two decidable logics that I call Z_1 and Z_1^* and which correspond naturally to interviews of G and G*. The corona $Z_1^* \backslash Z_1$ is not empty. In particular, I could show that Z_1^* proves the formula:

$$p \rightarrow \Box\Diamond p,$$

where p is a Σ_1 arithmetical proposition (and therefore here we need to restrain the substitution rules for $Z_1(^*)$).

Concerning our research into a purely arithmetical phenomenology of matter, we could say that there is both good and bad news.

[26] Note that Abramsky[1] and Vickers[64] have already modelled the notion of observable by similar propositions.

[27] In fact, the system V consisting of the axioms of G, accompanied by $p \rightarrow \Box p$, with the rules MP and NEC, is not only correct, but has been proved arithmetically complete for (Σ_1) provability. Even the *natural* system V* is complete for truth on these propositions[65].

- The good news: $p \to \Box\Diamond p$ is a modal formula, called B in the literature, allowing us to axiomatise the logic of quantum mechanical propositions. This comes from a result obtained by Robert Goldblatt (Goldblatt 1974). Like S4Grz, which formalises within modal transformation, the intuitionist logic of the subject, the system B, axiomatised by the axioms K, T, B with the rules MP and NEC, formalises, within a modal transformation, quantum logic. A miracle is operating here, because we start with an anti-symmetric logic of consciousness and arrive at a quasi-symmetric logic, as the formula B suggests.

- The bad news is that Z_1^*, like G^*, is not closed under NEC, and not even, in contrast to Z, closed under the monotony rule $\frac{p \to q}{\Box p \to \Box q}$, which characterises logics admitting Scott–Montague semantics. The miracle above is somewhat relative.

The bad news is not so bad as all that, but one must study more technical considerations in order to substantiate this proposition. What is surprising here is that the formula B is not demonstrated by Z_1. The "quantum" aspect, due to B, is strongly connected to a notion of third person plural. The *empirical* Plank's constant, which defines the level at which the quantum phenomena are incontrovertible, would also define the level of duplication where we survive as populations of machines.

The Z_1 logics allow us to formalise a certain number of natural questions on the *phenomenology* of isolated matter here. Do they violate Bell's inequalities? Which quantum logic must it be? Birckhoff and von Neumann's 1936 logic? Does it define a universal quantum machine? These are open problems, as is that of finitely axiomatising all Z logics.

We may hope to extract an algebraic semantics of Z_1^* logics in the form of a trellis of subspaces of a Hilbert space. In this case, we could use the unitary measure results to extract a Feynman formulation of quantum mechanics deduced purely from the discourse of a self-observing Universal Machine. We could therefore, start to hunt the white rabbit[28] and the flying pig ....

Having been independently verified by numerous people, I'm starting to hold the proof on The Reversal result as a given. I even find the strategy of the interview with the Universal Machine and its guardian angel as

[28] You can find an interesting discussion on the Net on different strategies for hunting the white rabbit in the form of metaphysics accepting the existence of all possible worlds at the address http://www.escribe.com/science/theory, or more latterly https://groups.google.com/forum/?hl=en#!forum/everything-list The earlier years of this discussion list has been summarised in [59].

natural. That is the difference between the logics of communicable propositions (based on G) and those containing true incommunicable propositions (based on G*) which permits us to clarify numerous obscure points in the philosophy of science and mind.

I am less sure however, of the pertinence of the present choice of Theaetetical definitions of knowledge and observation. Other choices are possible. In particular, we can reapply Theaetetus's idea and define a very weak intensional (modal) nuance of provability in studying the logic of provable, consistent and true propositions. What is truly astonishing, is that the Σ_1 restriction on S4Grz collapses all modalities: it only gives a propositional calculus.[29] But with this double iteration of Theaetetus's idea, we obtain a new "quantum" logic that proves the formula B (see the thesis) and which allows us to capture the notions of physical sensation or of "true" qualia.[30]

Again, we could be interested in all such logics with Σ_α restrictions, where α is one of Church and Kleene's constructive ordinals (equivalent to Cantor's constructive ordinals). Nevertheless, it is astonishing that a translation as "brutal" as the reversal argument also isolates rapidly an arithmetical interpretation of formula B. What is even more astonishing is that Theaetetus's idea at first leads to S4Grz where the Kripke semantics are antisymmetric. One might have feared a departure from physical logic where we would have also needed a symmetric Kripke semantics. Symmetry, which Maria Louisa Dalla Chiara, quantum logician from Florence says is welcome for a logic of physical propositions, preserves quantum idealism or computationalist subjectivism (or solipsism).

Note as well that the fact that the corona $\mathrm{Z}_1^* \setminus \mathrm{Z}_1$ is not empty allows us to explain why quantum logic semantics can serve to axiomatise the notion of physical *sensation* and qualia (c.f. also Bell J. L. 1986), since there are incommunicable observables (or measurables) (think of pleasure or pain).

The inspiration to use Plato's Theaetetus came from my reflections on dreams.[31] What struck me was the asymmetry existing between the states of dreaming and being awake: when you are awake, you can never be truly sure you are. By contrast, when dreaming you can sometimes perceive it

[29] This is incorrect since we must take account of the weakening of the substitution rule under the Σ_1 restriction. Thanks to Éric Vandenbussche for bringing this error to my attention. As $\mathrm{S4Grz}_1$ proves formula B, and is closed under the necessitation rule, it could constitute a new pointer towards an arithmetical quantum logic.

[30] Term used in cognitive science to designate the phenomenal contents of physical sensation.

[31] See the 1995 IRIDIA technical report[41]. It contains a very detailed chapter on the nature of dreams, as well as extracts from my dream diary. Since 1976, I noted my nightly dreams.

as such.[32] Nearly all the work was finished by 1986. I came to develop, however, the *arithmetisation* of the definitions of Theaetetus's knowledge at the Université Libre de Bruxelles, at IRIDIA more precisely, thanks to a national research project, which prompts a return to the main story of the thesis. At that moment, the "thesis" was only a hobby. I was looking for answers to the questions I had once asked. I really had the idea of one day writing articles or a book, but I no longer believed, since 1977 (see chapter 4), in the idea of mounting an entire *academic* thesis.

[32] We speak here of a lucid dream. Lucid dreams were put on an experimental footing by the parapsychologist Nearne, and then by the neurophysiologist-mathematician LaBerge in the 1980s[34].

Chapter 9

IRIDIA, *Mon Amour* (1987→...)

"Verhofstadt! Verhofstadt! ..."

I was deeply puzzled at what could possibly make my friend Professor Philippe Smets announce so vigorously the name of a well-known Flemish minister.

"Are you aware of *Verhofstadt*?" he presses on. "Do you still want to do modal logic?"

Euh ... Yes, but so what?

"Project Verhofstadt, two years, renewable perhaps to four, for fundamental research in Artificial Intelligence. Apply for it straight away, I need a modal logician at IRIDIA." "I would be able to do fundamental research?", I asked, captivated. "Absolutely", he told me. "You could even do a doctoral thesis."

The meeting between Philippe and I took place at a conference of the Belgian Society for logic and philosophy of science. Philippe was a doctor who worked on the problem of automating medical diagnosis. He specialised in medical statistics and was convinced of the irrelevance of statistics and probabilities to this area. He was interested in Dempster and Shafer's theory of "belief functions", to which he had contributed. He was persuaded of the relevance of logic to the formalising of aspects of this theory, and insisted that I come over to give a modal logic course at IRIDIA.

IRIDIA was the *Institut de Recherche Interdisciplinaire pour le Dévelopment de l'Intelligence Artificielle* (Institute for Interdisciplinary Research for the Development of Artificial Intelligence) that Philippe had established at Université Libre de Bruxelles. The institute had only existed

for a few months.

At the same time, the Head of the Biological Macromolecule Conformations Unit (UCMB), Shoshana Wodak, let me know that the central headquarters of Plant Genetic Systems (PGS) had turned-up the pressure on getting immediate and tangible results, so that my long-term project had been transformed into a short-term one. Michel was thinking of leaving UCMB and setting up his own company. He insisted that I accompany him and serve as a consultant for his company. He suggested that I be a consultant and at the same time benefit from Verhofstadt's fundamental research project.

I decided to leave PGS, abandoning ANIMA and refusing Michel's offer. As I am scrupulous, I sensed that the fundamental research I would be doing at IRIDIA would be incompatible with the consulting work. If this was to be the first time in my life that I was to gain finance for fundamental research, I did not wish to take the risk of being perturbed by overly practical questions that might distract me.

In January 1987, I started at IRIDIA. "Don't count on my thesis too soon, Philippe." "As you wish", Philippe told me. Philippe was *cool*, he had the *tao* of the effective boss. The researchers at IRIDIA had every freedom and were driven by their natural enthusiasm, fortified by discussions and regular brain-storming sessions. Each worked their own hours, and everybody evidently worked between 10 and 12 hours per day, the quality of the working environment being reflected in the quality of the results. Above all, IRIDIA was independent from all of the faculties, guaranteeing the necessary freedom for interdisciplinary cross-pollination. There was a coffee room and a ping-pong room; in truth IRIDIA was a tiny paradise for researchers. This was financed through European project funding, such as ESPRIT, or by national and international private projects.

I taught modal logic and Kripke semantics. I became the "Mr Kripke" of IRIDIA. I threw myself into it with all my heart. I also taught the rudiments of information theory and theoretical artificial intelligence following the work of Blum, Case and Smith .... This is why, after a rather long discussion lasting a year, Philippe Smets and I came to the conclusion that the modal logic KD formalised certain aspects of belief functions. This was something that Philippe would develop in detail with Natasha Aleshina, a Russian mathematician who worked in Amsterdam and who gained a similar result via an independent method (based on the work of Fattorosi and Barnaba who used KD for a modal approach to probabilities). This important point subsequently motivated me towards the weak version of "Theatetical" knowledge I use in chapter 5 of the Lille thesis (see also the preceding

chapter).

With the openness of spirit and good humour at IRIDIA, I finished up doing a presentation on the usage of modal logic in the theory of Gödellian self-reference, and I presented at last the rudiments of what I would call the "exact psychology of machines". The presentation was received very warmly. Philippe asked me if it was original. I explained to him the long and lively debate between researchers in the area concerning the relationship between Gödel's Incompleteness Theorem, machines and mind, starting from 1921 (Emil Post). Some, such as Lucas[1] think that Gödel's theorem refutes Mechanism. Others, such as Webb and myself, think that Gödel's theorem confirms Church's thesis, protecting Mechanism against numerous reductions in which it is usually circumscribed.

I added "All being said and done, I have perhaps an original result, *truly* original, which is either false or truly revolutionary (I can't help it)." "What's that?" I told him "a proof that if we are machines, there cannot be a universe. The appearance of a universe, or even universes, must be explained by the geometry of possible computations of possible machines, seen by these machines". Philippe asked me to present the "proof" to IRIDIA.

Paul Gochet, professor of logic at Liège university, had assisted for some time with my presentations on logic. Once more, he came to hear me. I told him of course, that he risked being deceived because I was not going to speak of logic, not directly in any case. I was planning on presenting the Movie Graph Argument and the RE paradox[2]. It consists of an old version of the Universal Dovetailer Argument. I didn't even get to touch on the RE paradox. My seminar was followed by a discussion that extended through the evening and part of the night. Gochet let me know that he was very interested and surprised, and suggested that I send a paper to the Cognitive Science Congress in Toulouse straight away. I had immense respect for Gochet, still more impressed by the discussion with my friend Dominique on his "Sketch of a Nominalist Theory of the Proposition". It was astonishing: Paul Gochet is a rare type of Belgian logician, passionate about analytical philosophy and a specialist on Quine. Analytical philosophers, more than anything at that time, dissolved antiseptically questions of the philosophy of mind. In fact, Paul Gochet understood that my approach led to a purely *mathematical* reformulation of the mind-body problem, and with the logic of self-reference, I had constructed a model wherein these

[1] This was in 1987, Penrose had not yet published "The Emperor's New Mind"[48], which re-launched the debate and spread it to the physics community.

[2] RE stands for "recursively enumerable", ie capable of being generated by a computer

questions possessed mathematical sense; arithmetical, even. He had heard, and his understanding was that I had put my finger on something.[3] Philippe Smets understood also that my stance was serious, but he was extremely sceptical as much over the results themselves as over their practical reach. He nevertheless allowed that there was in fact the makings of a thesis in this, and so came back to the crucial question "there will be only 36 Verhofstadt projects you know. It's now or never".

At Toulouse, where my paper was accepted, I was received very well and everyone asked me to publish it and do a thesis, so, it was back to Brussels. That became tiring. Truthfully, it was scary. I was going to tell Smets that I would indeed do a thesis, on the condition that I submitted it to Liège, or Louvain, or Toulouse: not to Brussels.

"Why?" "Let's say that within the Science Faculty, the Department of Mathematics is not very open to Gödel, and the Department of Computer Science is not very open to Artificial Intelligence." Philippe told me that he was not surprised: there was effectively, as might be expected, a cold war between IRIDIA and the Department of Computer Science. They were jealous and critical of IRIDIA at the same time. "But", he added, "they would have nothing against you, the thesis examiner panels at IRIDIA are rounded-out with foreign experts, it is sufficient to listen to you to see that you are making a valid argument. They cannot ridicule your work in public: what do you fear? That they will find a mistake in your proof? Just write your thesis, things will be fine. You are blowing things out of all proportion."

I had not spoken to him of X, nor of the ordeal I had endured. He wouldn't have understood. That concerned a final study project, nearly 20 years previously. Perhaps I may have exaggerated the difficulties.

So, I began to write this thesis. I detest writing quite as much as I enjoy debating with a listener. In order to be convincing, I need to see the eyes of those I address. I am aware that my propositions can strike others as paradoxical. I often interrupt myself to ask if there are any objections, which there always are. To clarify, I remove ambiguities. You show me when you know you understand, and so I move to the following step. I doubt the reader's patience for my writing. Those with a scientific background will not take seriously those passages which have a "philosophical air".[4] Those with a philosophical background will skip the technical sections. Who exactly *am* I addressing? I try to write for everyone, so I wrote my thesis between 1989

[3]Paul Gochet was the first to understand that I had transformed the mind-body problem into a search for a justification for the appearance of matter, and that this put doubt on the fundamental status of the physical sciences.

[4]Philosophy is grouped amongst the literary disciplines in French-speaking countries.

and 1992 as: "Conscience et Mécanisme",[41][5] circa 300 pages in length.

In the meantime, IRIDIA's personnel came and went. The great majority were Italians. There were also Chinese from the People's Republic of China, Chinese from Taiwan, Vietnamese, French, English, Germans and a minority of Belgians. By a type of magic that Philippe was to a great extent responsible for, the atmosphere there remained the same throughout: serene and enthusiastic.

From 1992 to 1994, I rewrote certain chapters and considerably improved the chapter on dreams and the Gödellian analysis of Descartes' *cogito* argument. I paid attention to the accuracy of the references. I described the routes and detours of my predecessors in the labyrinthine interference between Gödel and computationalist philosophy. I added numerous LISP programs which illustrated all the technical notions in the thesis, such as "amoebas", mirror programs, "planarians", and demonstrations of modal logic theorems using the Universal Dovetailer itself. The thesis grew from 300 pages to 750. I excused this to myself imagining that in reality it didn't matter whether I submitted or not. I started to resemble those poor researchers who seem incapable of finishing a thesis. One fine day, at the end of 1994, Philippe told me in no uncertain terms that the thesis was now finished and that I must submit.

"OK", I told him, "I'm going to submit to Louvain".

I had dreamt of submitting my thesis to the Catholic University at Louvain because that would get a serious label attached to my chapter on theology. Above all, I have always been well-received by logicians, as well as by mathematician/philosophers at Louvain, like Ladrière who, in particular was both philosopher and mathematician, but also Thierry Lucas and Marcel Crabbé. My only doubt was that submitting to Louvain would not be an easy path, but I was sure that it would engender a deep and interesting debate, and for me that was the only thing that mattered. The apparent contradiction between Mechanism and Catholicism was, in my opinion, due to the always-prevalent reductionist pre-Gödellian conception of machines.[6])

[5] Consciousness and Mechanism

[6] For an extreme example, see Jacques Arsac[3]. Evidently, computationalism, as I define it, is much more Platonic than Aristotelian. We are closer to Jean Trouillard[62], than to André Léonard[38], or Dominique Lambert[37]. I appreciate the appeal of a dialogue between scientists and theologians, but such a dialogue must not be used to minimise the importance of Platonism and to exclude eventual theologies (and theo-technologies) of an analytical or deductive nature. In the appendix on Church's thesis, I will suggest that CT rehabilitates Pythagorean philosophy, stripped of its more superstitious aspects. For a modern introduction to Pythagoras, see Dominic J. O'Meara[45] "Pythagoras Revived", Clarendon Press, Oxford, 1989. I can no longer resist citing O'Meara's beautiful little

"That would be a slap in the face for ULB", responded Philippe. I said that if I submitted to ULB, that would be rather a slap in the face for me, or worse: they will make me rewrite it ten times; they will keep knocking it back until there is nothing left.

Philippe let me know that he was fed up with my "paranoia". And I could see that I gave the appearance of a perfect paranoid. How might I have gone against his wishes? *Was* I paranoid? Intellectually I could well believe it, but in my guts, I felt I was being sent straight to the abattoir.

book, "Plotinus: An Introduction to the Enneads"[46], Clarendon Press 1995.

Chapter 10

Darker Than You Think [II] (1995–1998)

> *Le refus de communication direct est l'arm absolue de pervers.*
> *The refusal of direct communication is the ultimate weapon of the perverse.*
>
> Marie-France Hirigoyen

Early in 1995, I submitted my thesis to Brussels. I didn't want to inconvenience Philippe any further, and after all, it was thanks to him and IRIDIA, an entity of ULB, that I had written this work. Also, the Head of the Mathematics department, with whom I had had good relations during my studies, and to whom I had shown a copy of my thesis, had me practically reassured. Having brought him up to date over the problematic end-of-studies dissertation, he made me see that it all happened more than 20 years ago, that the wind had changed; that artificial intelligence was taken seriously now, and that nobody from the department would take the risk of looking ridiculous in front of foreign experts; that I had nothing to fear, that he would personally watch over it, etc.

For reasons of academic courtesy, he strongly suggested I offer a copy of the thesis to X, and ask him to be the "official" sponsor (because "that's how it is done") whereas Philippe Smets would in fact be my real supervising sponsor.

And so I gave the thesis and the department head's proposal to X: he accepted the copy without as much as blinking. "I will be in touch", he finished, as he showed me the door. Four days later, by email, he stated

that he refused to be my thesis sponsor. What a pity! So much the better! Phew!

A well-intentioned person suggested I give a copy to a certain Y (say). He had a good reputation, he gave courses on the history of mathematics, and I remember with great fondness his course on number theory. Good memories.

One small niggle. Amongst friends, over a glass, I confide that I believe Y was communist in the 70s.

1973–77 were my student years. I didn't have a reputation for being truly *communist* at that time. Worse, I belonged to an Orwellian left that never believed in the success of *La Révolution*. I owe this to several factors. Simon Leys is the brother of one my grandmother's associates at Éditions Larcier, and offered me all these oh-so-lucid books on China.[1] My father was a military man, and I would say he had the correct viewpoint in my eyes, and was something of a living counterexample to the prevailing anti-militaristic discourse. Above all, in 1968, I was too little, before my encounter with Watson.

My friends and I fell about laughing and shouting "paranoid! paranoid!". "OK, OK, my friends, I'll give a copy to Y". Sigh.

Fate intervened, and had it that I encounter Y in a parking lot.
"M. Y, I would like to give you a copy of my thesis...".
"Er... that is... no thanks, ...".
"Why?", I spontaneously asked.
"Its that ... X told me it was no good".
"... and you don't want to make up your own mind?", I exclaimed, just as spontaneously.

He left, with a vexed air. "Gosh!", I thought.

Philippe was a little astonished and disappointed with X and Y's reaction, and was astonished at my relief. I was relieved to see X get out of the way.

We decided to follow the typical procedure for this situation, which consisted of proposing a list of possible professors to make up the panel, with foreign experts (outside the Mathematics Department in the Faculty). Ten or so ULB professors (physicists, engineers, biologists, etc.) plus five or so external professors.

Result: outcry and hysterics in the Mathematics and Computer Science departments. I don't understand. Long months of suspense. They told me to wait, that things would calm down. A problem with my thesis? "Nothing

[1] "The Chairman's new clothes: Mao and the Cultural Revolution", Simon Leys

to see", they told me, "we are being careful to not cause a wrong reaction."

Then the calm, and finally, the good news: we had been granted an open field to make up the panel.

Then, the other good news, at least according to Philippe. X and Y asked to be members of the panel."You see, they are interested in your work, the wind has changed, they know that your work will pass, and they want to be in on that". I will not try to explain my reaction to this. I was speechless.

After that, I received a sort of official notification that my thesis had been effectively accepted and that it remained only to fix the date of the Viva. Reassured finally, I committed the error of sending my thesis to 50 impatient souls that I had promised to send "my thesis" one fine day, and had therefore saved their addresses. Perhaps if I hadn't promised, I would never have sent them ... sending out 50 copies of 750 pages is quite a job!

I became increasingly nervous as I still did not know the makeup of the examination panel.

26/9/1995. 10 o'clock in the morning, my place. Telephone rings. The Science Faculty Mathematics department secretary lets me know the panel makeup. The department Head is leading the panel, X is the sponsor and the members: Y, Philippe and three other experts (from ULB, and designated by the panel leader). And nobody else? They didn't choose a single person from the 15 experts proposed by Philippe. That stinks of foul play.

At the same time, I was slightly reassured, since during a private defence between the departmental head and Philippe, I could not see any mention of how X and Y might demolish me (outside of finding a "genuine" mistake in my work for sure, but after a year, it seemed my work was the last concern).

"What are you afraid of now? Are you angry that they will take credit for the success of your work? Is that it? Hey—that's how glory is, old friend!", Philippe told me.

I wasn't anything like at ease, but the presence of the departmental head had reassured me.

27/9/1995. 2 o'clock in the afternoon. Philippe had just gone to the United States, and the telephone rang in my office at IRIDIA. A panicked voice. The same department secretary, I can hear raised voices in the background. She tells me that there has been a mistake in the panel makeup: Y is the panel lead, X the sponsor, etc. The department Head had disappeared, with no means of telephone contact with the panel! At that moment, I realised it was finished. I felt an infinite rage boiling up inside me, so I sent a somewhat dry email to X along the lines of "why?".

On Philippe's return from the United States, I wasn't alone. The three

of us explained to him that without any doubt, the whole manoeuvre was as obvious as house paté. There would be no Viva, private or public, and we insisted he in fact refuse to go to this meeting revisiting the acceptability of my thesis, because at that time and place it would be judged unacceptable. Philippe became angry: you would never see this at ULB, moreover, he was going to convince them of the necessity of enlarging the panel in light of the vast character of the work, etc. Philippe still believed that I feared that they were stealing my "glory", or acting in bad faith, preventing me from having a grade. That was the worst that he could possibly imagine.

"You're dreaming, Philippe! For more than 20 years they *never* let me say so much as a word. They're not about to start today. You know very well they ridicule artificial intelligence, they misrepresent engineers and philosophers and they have a wild hatred towards IRIDIA" I explained to him.

"You exaggerate, and you will see". Philippe was trapped by his confidence and optimism, qualities that made IRIDIA such a good environment. He was unable to entertain that they might refuse the thesis without listening to me at least once, in private, such is the practice at ULB. And so, Philippe went to the meeting. He came back as white as a sheet. The work was judged unacceptable, without any possibility of appeal.

Relief at not having to deal with them, even though they lost in a stroke their scientific credibility in my eyes; their *rationalist* credibility even. Relieved that having written the thesis, it would not be attacked. Relieved, finally, that they would have to, in contrast to 1977, sign their opt-out forfeit note. In effect, the process required them to produce a report of unacceptability. I harboured no illusions. The report was a pure formality and the process required only that the thesis title appear along with the word "unacceptable". Three weeks later, I received this report, which consisted of a single page, dated and signed, which effectively mentioned the thesis title, and the word "unacceptable", and little else. The message was that everything was correct, according to the experts, but that it didn't contain any original results!

You might be gaining the impression that they didn't want to risk any oral confrontation with me.

I had been judged without having been heard (or read), on two occasions therefore, by essentially the same person, his friends, and twenty years apart.

As for the question I tried to put in 1963, and the embryonic answer I tried to share since 1971, that started to mature.

Fatigue and exhaustion.

Two professors strongly encouraged me to present my thesis under their

supervision. Professor Paul Gochet from Liège University and Professor Jean-Paul Delahaye from Lille University. This was both an intellectual and a moral comfort. Some years previously, M. Gochet sent my papers to M. Delahaye, who invited me to Lille to discuss my approach. I had accepted the supervision by Jean-Paul Delahaye. He suggested to be as clear as possible, and pressed me to put into relief the most original part, the most surprising part without doubt: The Reversal.

Enthusiastically, Jean-Paul Delahaye published a paper on my work in the review "Pour La Science". With the defense![19]

For quasi-administrative reasons, one must undertake a two year diploma of advanced studies in computer science, so I had to wait two years before defending my thesis at Lille on 2nd of June, 1998.

On seeing so many Belgian number plates arrive in the Lille University car park, Jean-Paul asked me, a little worried, whether some of these might be my "opponents". Frankly, I smiled, it was more a case of my friends. Jean-Paul couldn't have known anything since I am here telling this story for the first time. And all of that is now well past. I wondered how many questions Professor Paul Gochet would ask me; they all turned out interesting, and he stopped at six. It gave me great pleasure to answer him, as it did to all the questions asked by the panel. The panel chairperson commented on my masterful presentation, the panel members warmly congratulated me, and I thanked them. I was happy for them that they got to hear me.

Brussels, 19th May, 2000.

Bibliography

[1] S. Abramsky. *Domain Theory and the Logic of Observable Properties.* PhD thesis, University of London, 1987.

[2] Claude Allègre. *The Defeat of Plato.* Fayard, Paris, 1995.

[3] Jacques Arsac. *Les Machines à penser, des ordinateurs et des hommes.* Seuil, 1987.

[4] Alain Aspect, Phillipe Grangier, and Gérard Roger. Experimental realization of Einstein-Podolsky-Rosen gedankenexperiment; a new violation of Bell's inequalities. *Phys. Rev. Lett.*, 49:91, 1982.

[5] J. S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1964.

[6] Charles H. Bennett. Logical depth and physical complexity. In R. Harkin, editor, *The Universal Turing Machine: A Half Century Survey*, volume 1, pages 227–258. Oxford UP, 1988.

[7] S.J. Blackmore. *In search of the light: The adventures of a parapsychologist.* Prometheus Books, 1996.

[8] George Boolos. *The Unprovability of Consistency.* Cambridge UP, London, 1979.

[9] George Boolos. *The Logic of Provability.* Cambridge UP, Cambridge, 1993.

[10] J.L. Borges and A. Hurley. *Collected fictions.* Viking New York, 1998.

[11] Ralph Buchsbaum. *Animals without Backbones: 1.* Pelican, 1938.

[12] M. Burnyeat. Socrate et le jury: de quelques aspects paradoxaux de la distinction platonicienne entre connaissance et opinion vraie. In

M. Canto-Sperber, editor, *Les paradoxes de la connaissance, essais sur le Ménon de Platon*, pages 237–251. Odiles Jacob, 1991.

[13] C. Butterworth. *Averroes' Middle Commentary on Aristotle's Poetics*. St. Augustines Press, 1999.

[14] B.F. Chellas. *Modal Logic: an Introduction*. Cambridge university press, 1980.

[15] Claude Cohen-Tannoudji, Bernard Diu, and Franck Laloë. *Quantum Mechanics*. Wiley, 1977.

[16] A. Connes, A. Lichnerowicz, and M.P. Schützenberger. *Triangle of thoughts*. Amer Mathematical Society, 2001.

[17] M. Davis, editor. *The Undecidable*. Raven Press, Hewlett, NY, 1965.

[18] Louis de Broglie. *La théorie de la mesure an mécanique quantique ondulatoire*. Gauthier-Villar, Paris, 1957.

[19] Jean-Paul Delahaye. Le monde des machines. *Pour La Science*, 243:100–104, 1998.

[20] D.C. Dennett and D. Hofstadter, editors. *The Mind's I*. Basic Books, NY, 1981.

[21] Bernard d'Espagnat. *Conceptions de la physique contemporaine; les interprétations de la mécanique quantique et de la mesure*. Hermann, Paris, 1965.

[22] Bernard d'Espagnat. *Conceptual Foundations of Quantum Mechanics*. Addison Wesley, 2nd edition, 1976.

[23] David Deutsch. Quantum theory, the Church-Turing principle and the universal quantum computer. *Proc. Royal Soc. London Series A*, 400:97—-117, 1985.

[24] A. Einstein, B. Podolsky, and N. Rosen. Can quantum mechanical description of physical reality be considered complete. *Phys. Rev.*, 47:777, 1935.

[25] Richard P. Feynman. imulating physics with computers. *International Journal of Theoretical Physics*, 21:467——488, 1982.

[26] Daniel Galouye. *Simulacron 3*. Phoenix Pick, Rockville, MD, 1965.

[27] Jean-Louis Gardiè. *Essai sur la logique des modalités*. Presses Universitaires de France, 1979.

[28] Kurt Gödel. Eine interpretation des intuitio9nistischen aussagenkalküls. *Ergebnisse eines Mathematischen Kolloquiums*, 4:39–40, 1933.

[29] Kurt Gödel. Remarks before the Princeton bicentiennial conference on problems in mathematics. In Davis [17], pages 84–88.

[30] J. Haugeland. *Artificial intelligence: The very idea*. MIT press, 1989.

[31] William Hayes. *The Genetics of Bacteria and Their Viruses*. Blackwell Scientific Publications, 2nd edition, 1970.

[32] S. C. Kleene. *Introduction to Metamathematics*. North-Holland, 1952.

[33] A. N. Kolmogorov. Zur deutung der intuitistischen logik. *Math. Zeitschr.*, 35:58–65, 1932.

[34] Steven LaBerge. *Lucid Dreaming*. J. P. Tarcher, Los Angeles, CA, 1985.

[35] J. Ladrière. *Les limitations internes des formalismes*. E. Nauwelaerts, Louvain, 1957.

[36] Jacques Lafitte. *Réflexion sur la science des machines*. Vrin, 1972.

[37] Dominique Lambert. *Science at théologie, les figures d'un dialogue*. Press Universitaires, Namur, 1999.

[38] André Léonard. *Foi et Philosophe, guide pour un discernment chrétien*. Cultur et Vérité, Namur, 1971.

[39] B. Marchal. Informatique théorique et philosophie de l'esprit. In *Actes du 3eme colloque international de l'ARC*, pages 193–227, 1988.

[40] Bruno Marchal. L'elaboratore è un grafo. *L'insegnamento della matematica e delle scienze integrate*, page 43, March 1983.

[41] Bruno Marchal. Conscience et mécanisme. Technical report, Université Libre de Bruxelles, 1994. http://iridia.ulb.ac.be/~marchal/bxlthesis/consciencemecanisme.html
accessed 11th October 2012.

[42] Bruno Marchal. *Calculabilité, Physique et Cognition*. PhD thesis, L'Université des Sciences et Technologies de Lille, 1998. http://iridia.ulb.ac.be/ marchal/publications.html.

[43] Tim Maudlin. Computation and consciousness. *J. Philosophy*, 86:407–432, 1989.

[44] J. C. McKinsey and A. Tarski. Some theorems about the sentential calculi of Lewis and Hayting. *J. Symbolic Logic*, 13:1–15, 1948.

[45] Dominic J. O'Meara. *Pythagoras Revived*. Clarendon, Oxford, 1989.

[46] Dominic J. O'Meara. *Plotinus: and Introduction to the Enneads*. Clarendon, Oxford, 1995.

[47] Linus Pauling. *General Chemistry*. W.H. Freeman, 1970.

[48] Roger Penrose. *The Emperor's New Mind: Concerning Computers, Minds and The Laws of Physics*. Oxford UP, Oxford, 1989.

[49] Karl R. Popper. *The Open Society and Its Enemies*. Hutchinson, London, 1950.

[50] Emil Post. Recursively enumerable sets of positive integers and their decision problems. *American Mathematical Society*, 50:284–316, 1944.

[51] J. Richelle. Contribution à l'étude théorique de certains aspects de la régulation de l'immunité du bactériophage tempéré λ. Mémoire de license en science chimiques, Université Libre de Bruxelles, 1974.

[52] F. Rivenc. *Introduction à la logique*. Payot & Rivages, 2003.

[53] Bernard Roques. *"La dangerosité des drogues", Rapport au secrétariat d'État à la Santé*. Edition Odile Jacob, 1999.

[54] Rudy Rucker. *Infinity and the Mind*. Harvester, Brighton, Sussex, UK., 1982.

[55] Ehud Shapiro. *Algorithmic Program Debugging*. MIT Press, Cambridge, MA, 1983.

[56] P. Smoryński. *Self-Reference and Modal Logic*. Springer, New York, 1985.

[57] R.M. Smullyan. *Forever Undecided*. Knopf, 2012.

[58] R. M. Solovay. Self-reference and modal logic. *J. Mathematics*, 25:287–304, 1976.

[59] Russell Standish. *Theory of Nothing*. Booksurge, 2006.

[60] J.H. Taylor. *Selected Papers on Molecular Genetics*. Academic Press, New York & London, 1965.

[61] J. P. Thibaudat. De l'autre côté du rideau rouge. *Libération*, Jul 1989.

[62] Jean Trouillard. *L'Un et l'Âme selon Proclos*. Belles Lettres, Paris, 1972.

[63] Alan Turing. On computable numbers with an application to the entscheidungsproblem. *Proc. London Math. Soc.*, 42:230–265, 1936.

[64] S. J. Vickers. *Topology via Logic*, volume 5 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge UP, Cambridge, 1989.

[65] A. Visser. *Aspects of Diagonalization and Provability*. PhD thesis, University of Utrecht, 1985.

[66] J. C. Webb. *Mechanism, Mentalism and Metamathematics: An essay on Finitism*. Reidel, Dordrecht, 1980.

Index