Undecidability everywhere?

K. Svozil

Institut für Theoretische Physik

University of Technology Vienna

Wiedner Hauptstraße 8-10/136

A-1040 Vienna, Austria

e-mail: svozil@tph.tuwien.ac.at

www: http://tph.tuwien.ac.at/~svozil

casti.tex

Abstract

We discuss the question of if and how undecidability might be translatable into physics, in particular with respect to prediction and description, as well as to complementarity games.

1 Physics after the incompleteness theorems

There is incompleteness in mathematics [?, ?, ?, ?, ?, ?, ?, ?]. That means that there does not exist any reasonable (consistent) finite formal system from which all mathematical truth is derivable. And there exists a "huge" number [?] of mathematical assertions (e.g., the continuum hypothesis, the axiom of choice) which are independent of any particular formal system. That is, they as well as their negations are compatible with the formal system.

Can such formal incompleteness be translated into physics or the natural sciences in general? Is there some question about the nature of things which is provable unknowable for rational thought? Is it conceivable that the natural phenomena, even if they occur deterministically, do not allow their complete description?

Of course it is! Suppose there exists free will. Suppose further that an observer could predict the future. Then this observer could freely decide to counteract in such a way as to invalidate that prediction. Hence, in order to avoid paradoxes, one has either to abandon free will or accept that complete prediction is impossible.

The above argument may appear suspiciously informal. Yet, it makes use of the diagonalization technique, which is one of the royal roads to a constructive, rational understanding of undecidability in the formal sciences. What Gödel and others did was to encode the argument in a language acceptable to their area of research. To translate and bring similar issues into mainstream natural

science is, at least in the author's opinion, the agenda of the present concern on rational limits to science.

Before discussing these questions further, we should first clarify the terms we use. Under a *physical phenomenon* we shall understand an event, which is (irreversibly) observed. A typical physical phenomenon consists of a click in a particle detector: there can be a click or there can be no click. This yes-no scheme is experimental physics in-a-nutshell (at least according to a theoretician). From this type of elementary observation, all of our physical evidence is accumulated.

Then there are *physical theories.* Physical theories purport to relate to the physical phenomena. At face value, they consist of phenomena themselves: as observers, we would not be able to know about theories if we would not observe their representation or code. A typical code of, say, the theory of electrodynamics, consists of letters and symbols printed in a book on electrodynamics. Their reading corresponds to an act of observation.

Why should anyone bother about the intrinsic representation of physical entities such as observations and theories? Because this issue is crucial for an understanding of undecidability. Gödel, for instance, proved his incompleteness theorems as he succeeded to properly code a (generic) theory about arithmetic *within* arithmetic. (To put things in the right historical perspective: the Viennese Circle of positivists, most notably Carnap, provided the basis for such coding techniques.)

At one point of the argument, we have to confront ourselves with the question, "is there a physical correspondent to the notion of inconsistency; that is, to a logical contradiction [**?**]?" Can a particle, for example, be here and somewhere else (≡ not here) [**?**]? On the phenomenological level, the answer is no. To put it pointedly: there is no such thing as an inconsistent phenomenon. In a yes-no experiment which can have two possible outcomes, only one of these outcomes will actually be measured. In contradistinction, a *theoretical description* might allow the consistent "existence" of mutually exclusive states if it is indeterministic (probabilistic). We shall come back to these issues later.

Undecidability occurs in indeterministic as well as in mechanistic systems. The term *indeterminism* here stands for any process which cannot be described finitely and causally. As a metaphor, we may say that in indeterministic physical systems, "God plays dice." By definition, indeterminism implies undecidability. If there is no cause, there cannot be any predictable effect. That is the whole story. Period.

Let us be more specific and consider two examples: Firstly, in quantum mechanics, the prevalent probabilistic interpretation of the wave function pretends that it is a complete description. Single outcomes cannot be deterministically accounted for. This is the quantum dice. Secondly, in the scenario of "deterministic chaos," the (Martin-Löf/Solovay/Chaitin) randomness of "almost all" initial values represented by elements of the classical mechanical continuum is successively recovered during the time evolution—bit after bit. Therefore, if

one believes in the quantum dice and in the physical relevance of the classical continuum, then, by definition, there is undecidability in physics.

Why cannot we stop here, sit back and relax? We have just encountered the fact that present-day physical theories contain indeterministic features which evade any complete prediction. Why is this not the end of the story? The trouble is that we shall never be sure that the probabilistic interpretation of the wave function is complete, nor do we know whether the classical continuum is appropriate [**?**]. There might be a "secret arena" hidden to us momentarily, in which everything can be deterministically accounted for. If we relax now and uncritically accept indeterminism as a matter of unquestionable fact, we may be heading for trouble. Indeterminism, as it is conceived by the physics community at the *fin de siécle (millénaire)*, might be a degenerative research program.

Therefore, it seems not inappropriate to try to re-interpret physical indeterminism constructively. In doing so, it is necessary to study undecidability and incompleteness in systems which are mechanistic in more detail. By *mechanistic* we mean that they are finitely describable and causal in all of their aspects. (In what follows, the terms *mechanistic, computable* and *recursive* are synonyms.) In mechanistic systems, every effect has a cause. But, one may doubt, if everything has a cause, everything can be deterministically accounted for, computed and predicted. In such a scenario, how does one recover incompleteness?

The answer to this question derives from a most important epistemological issue: Although in principle every effect may have a cause, such causes might

not be knowable by intrinsic observers. We are introducing an inside-outside distinction here. *Intrinsic observers* are embedded in the system they observe—their "Cartesian prison" [?, ?, ?, ?, ?, ?, ?, ?]. They cannot step outside. Intrinsic observers are bound to observations which are intrinsically operational [?]. They can only refer to intrinsic entities [?] (cf. the "virtual backflow" [?]). Their theories must be intrinsically codable (cf. above).

Rather then attempting a formalization of intrinsic perception (cf. [?]), let us consider a metaphor, or rather a nightmarish virtual reality game or a Zen *koan:* Suppose that we are thrown into a prison (or a lunatic asylum, who knows?) without any explanation. In this prison we see persons vigorously talking to each other. Yet we do neither understand their language nor the reasons, rules and laws of that establishment.—In a methaphorical sense, the world might be perceived as such a prison, and science might been one attempt amongst others to make sense out of the situation.

But let us continue with determinism. The way it was defined, a mechanistic physical (dynamical) system corresponds one-to-one to a process of computation. This computation can, for instance be implemented on a universal Cellular Automaton, a universal Turing machine or any other universal computer. It, in turn, corresponds one-to-one to a formal system of logic. With these two identifications, namely *mechanistic dynamical system* $\equiv$ *computation* $\equiv$ *formal system*, we bridge the gap to formal undecidability. In principle, the term *system* could stand for any of these three entities [?, ?].

We should be quite clearly aware of the fact that there is no other possible formalization of undecidability besides recursive function theory and formal logic. If one resents the idea of logical or of computer science terminology creeping into physics, then there is no room for this issues. Undecidability in physics marks the integration of yet another abstract science—recursion theory—into physics.

We have set the stage now. Let us recapitulate: we would like to consider mechanistic physical systems. Intrinsic observers are embedded therein. These intrinsic observers register physical phenomena which are operational. Moreover, they develop theories which are intrinsically codable. Our aim is to figure out whether or not, under these constraints, certain physical phenomena and theoretical propositions become undecidable. In doing so, we have to translate the pandemonium of recursion theoretical undecidability into physics. Our translation guide will be the equivalence between mechanistic physical systems, computations and formal systems.

2 Prediction and description

Gödel himself did not believe in any physical relevance of the incompleteness theorems, in particular not for quantum mechanics [?]. One might speculate that he had been brainwashed by Einstein, who was bitterly opposed to the Copenhagen interpretation of quantum mechanics. Einstein thought that quan-

tum mechanics and the Copenhagen interpretation thereof was a degenerative research program [**?**, **?**]. Einstein's *dictum* "God does not play dice" has become a legend.

And yet, there is a straightforward extension of formal incompleteness to physics. It is based on Turing's proof that certain propositions about universal computers—basically modeled to mimic elementary mechanical paper-and-pencil operations on a sheet of paper—are undecidable. In particular, it is impossible to predict whether or not a particular computation task on universal computers will eventually terminate (and will output a particular result). Therefore, if we construct a physical device capable of universal computation, there are some propositions about the future of this system which are provable undecidable.

Let us be more specific and (algorithmically) prove the statement above. It is often referred to as the "halting theorem" or the "recursive undecidability of the halting problem." We are using a technique of *diagonalization,* which was pioneered by Cantor in a proof of the undenumerability of the reals. This technique is the most useful tool in exploring the undecidable.

The strategy of diagonalization (and related techniques) is to assume a statement—whose existence should be disproved—and, by trivial manipulations, derive a paradox, a contradiction. The only consistent way to avoid this paradox is to abandon the original statement. For the purpose of a formal proof, any paradox can in principle be exploited, as long as it is codable into formal enti-

ties. Gödel [?] as well as Turing [?] used "the liar" [?] for their incompleteness theorems. Gödel was well aware of the fact that almost any classical paradox might do as well. (Readers not interested in the details of the proof may skip the entire next section.)

We shall prove the recursive unsolvability of the halting problem algorithmically. That is, we shall use informal terminology which—by the Church-Turing thesis [?, ?]—is supposed to correpond to formal expressions.

Assume, for the moment, that there is an mechanistic way to foresee a particular aspect of the future of an arbitrary computation. Namely, whether or not the computation will terminate. Or, if it outputs a string of symbols and then terminates. Consider an arbitrary algorithm $B(x)$ whose input is a string of symbols x. Assume that there exists a "predictor" `PREDICT` which is able to decide whether B terminates on x or not. Using the predictor `PREDICT`$(B(x))$, we shall construct another computing agent A, which has as input any effective program B and which proceeds as follows: Upon reading the program B as input, A makes a copy of it. This can be readily achieved, since the program B is presented to A in some encoded form, i.e., as a string of symbols. In the next step, the agent uses the code of B as input string for B itself; i.e., A forms $B(B)$. The agent now hands $B(B)$ over to the prediction subroutine `PREDICT`. Then, A proceeds as follows: if `PREDICT`$(B(B))$ decides that $B(B)$ halts, then agent A does not halt; this can for instance be realized by an infinite `DO`-loop; if `PREDICT`$(B(B))$ decides that $B(B)$ does *not* halt, then A halts. (This is the

diagonalization step.) We shall now confront the agent A with a paradoxical task by choosing A's own code as input for itself. Notice that B is arbitrary and has not been specified yet. The deterministic agent A is representable by an algorithm with code of A. Therefore, it is possible to substitute A for B. Assume that classically A is restricted to classical bits of information. Then, whenever $A(A)$ halts, `PREDICT`$(A(A))$ forces $A(A)$ not to halt. Conversely, whenever $A(A)$ does not halt, then `PREDICT`$(A(A))$ steers $A(A)$ into the halting mode. In both cases one arrives at a complete contradiction. Classically, this contradiction can only be consistently avoided by assuming the nonexistence of A and, since the only nontrivial feature of A is the use of the predictor subroutine `PREDICT`, the impossibility of any such universal predictor.

Notice that the above argument is nothing but a rephasing of the informal argument against free will or complete predictability given at the beginning! Popper [**?**] considered these issues already in the forties. More sophisticated models have been put forward by Wolfram [**?**], Moore [**?**] and da Costa and Doria [**?**]. These approaches essentially embed a universal computer (or equivalent systems of Diophantine equations) into a classical physical structure such as a field. The system is assumed to be infinite to allow for infinite tape or its equivalent. Then undecidability follows, for instance, from the recursive unsolvability of the halting problem.

In short: reasonable (consistent) theories predicting the future behavior of arbitrary mechanistic physical systems are impossible. So, if one beliefs in the

physical relevance of the model of universal computers, then no physical theory can predict all the physical phenomenology. In particular, there are certain physical prediction tasks which are undecidable.

But what if one insists that any computation should remain finite? Then, in principle, it would be possible to construct a predictor, which would just have to simulate the system long and fast enough to complete the prediction. Could such a prediction take a sufficiently short time in order to be useful? And what if a finite predictor tries to predict itself? These questions get us into quantitative issues, which are more involved. We shall attack them next.

The busy beaver function [**?**, **?**] addresses the following question: given a system whose size of description is finite; more precisely; less than or equal to n bits long. What is the biggest number $\Sigma(n)$ which can in principle be produced by such a system before halting (or, alternatively, before recurring to the system's original state)?

A related question is: what is the upper bound of running time (or, alternatively, recurrence time) of a program of length n bits before terminating? An answer to that question gives us a feeling of how long we have to wait for the most time-consuming program of length n bits to hold. That, of course, is a worst-case scenario. Many programs of length n bits will have halted before the maximal halting time. Let us denote it by `TMAX`. We could figure out that knowledge of `TMAX` "solves" the halting problem quantitatively. Because if we knew that maximal halting time, then for an arbitrary program of n bits, we

would have to wait just a little bit longer than $\texttt{TMAX}(n)$. If it would still run, then we could be sure that it would run forever. Otherwise it would have halted. In this sense, knowledge of TMAX is equivalent to possessing a perfect predictor PREDICT. Since the latter one does not exist, we may expect that TMAX cannot be constructive function easy to work with.

Indeed, Chaitin has proven [?, ?, ?, ?, ?] that $\texttt{TMAX}(n) = \Sigma(n + O(1))$ is the minimum time at which all programs of size smaller than or equal to n bits which halt have done so. For large values of n, $\Sigma(n)$ grows faster than any computable function of n; more precisely, let f be an arbitrary computable function, then there exists a positive integer k such that $\Sigma(n) > f(n)$ for all $n > k$.

You can virtually see that any system trying to evaluate the busy beaver function "blows itself up." Originally, T. Rado [?] asked how many 1's a Turing machine with n possible states and an empty input tape could print on that tape before halting. The first values of the Turing busy beaver $\Sigma_T(x)$ are finite and are known [?, ?]: $\Sigma_T(1) = 1$, $\Sigma_T(2) = 4$, $\Sigma_T(3) = 6$, $\Sigma_T(4) = 13$, $\Sigma_T(5) \geq 1915$, $\Sigma_T(7) \geq 22961$, $\Sigma_T(8) \geq 3 \cdot (7 \cdot 3^{92} - 1)/2$.

What does all this mean for physics? One consequence is that, for mechanistic (but unbounded) systems describable by n bits, the recurrence time grows faster than any computable number of n. It is therefore uncomputable and thus impossible to predict.

Any causal prediction requires a theory of the system which one wants to

predict. In the intrinsic observer scenario described above, there is no way to cut out or separare the observer from the system. We have to deal with self-description.

Can observers embedded in a system ever hope for a complete theory or self-description? Let us rephrase the question. Is it possible for a system to contain a "blueprint," a complete representation, of itself? This issue has been raised by von Neumann in his investigation of self-reproducing automata. Indeed, von Neumann showed that—provided such a "blueprint" exists—a (universal) automaton can reconstruct a perfect replica of itself [?, ?, ?].

To avoid confusion, it should be noted that it is never possible to have a finite description with itself as proper part. The trick is to employ *representations* or *names* of objects, whose code can be smaller than the objects themselves and can indeed be contained in that object (cf. [?], p. 165). Gödel's first incompleteness theorem is such an example. Any book of electromagnetism is another.

A completely different issue is how such a theoretical self-description is obtained. Here we have to make a distinction. As in the above case, a complete theory or self-description might be obtained *passively* from some "intuition," "God" or "oracle." (Of course, one could never be sure that it is the right one.) But it is generally impossible for an intrinsic observer to *actively* examine the own system and thereby to construct a complete theory. One reason for this is self-interference and complementarity, as described below.

As a comfort to those who conceive the "Cartesian prison" as the source of all problems, one could cite a nice theorem by Gold [**?**]. It is sometimes referred to as the recursive undecidability of the rule inference problem: For any mechanistic intelligence/agent A, there exists a total recursive function f such that A does not infer f. In more physical terms, there is no systematic way of finding a deterministic law from the (input/)output analysis of a mechanistic physical system.

An informal way to (algorithmically) prove Gold's theorem uses the halting theorem. Suppose that it would indeed be possible to derive laws systematically. Let us call this agent or computable function `RULE`. `RULE` would have to watch the behavior of the system it analyzes. But any complete analysis would require the observation of `TMAX`(n), where n is the minimal description size of the system. Since `TMAX`(n) grows faster than any computable function of n, `RULE` cannot be computable.

So, even if we would be in a "God-like" position and would be disentangled and freed from the observed system, we would have to cope with the fact that there is no systematic way of deriving causal laws.—Indeed, we may safely state that, except for the most elementary phenomena, deriving causal laws remains a rare art!

Of what use is a complete theory? Is it possible for an observer in a finite amount of time to predict the own state completely?

An intuitive understanding of the impossibility of complete self-comprehension

can be obtained by considering a variant of Zeno's paradox of Achilles and the Tortoise/Hector (called "paradox of Tristram Shandy" by Popper Popper [?]): In order to predict oneself completely, one has to predict oneself predicting oneself completely, one has to predict oneself predicting oneself predicting oneself completely, one has to ..."

3 Complementarity Games

The Hinduistic notion of Maya suggests that the world of senses is illusory, that observations are distractive. Plato's cage metaphor emphasizes the distinction between objects and what we may be able to observe from these objects. Some day, complementarity might be perceived as a variation of this ancient theme.

There has been hardly any feature of quantum mechanics which has given rise to as many fruitless speculations as *complementarity.* Intuitively, complementarity states that it is impossible to (irreversibly) observe certain observables simultaneously with arbitrary accuracy. The more precisely one of these observables is measured, the less precisely can be the measurement of other—complementary—observables. Typical examples of complementary observables are position/momentum (velocity), angular momentum in the x/y/z direction, and particle number/phase [?, ?].

The intuition (if intuition makes any sense in the quantum domain) behind this feature is that the act of (irreversible) observation of a physical sys-

tem causes a loss of information by (irreversibly) interfering with the system. Thereby, the possibility to measure other aspects of the system is destroyed.

Well, this is not the whole story. Indeed, there is reason to believe that—at least up to a certain magnitude of complexity—any measurement can be "undone" by a proper reconstruction of the wave-function. A necessary condition for this to happen is that *all* information about the original measurement is lost. Schrödinger, the creator of wave mechanics, liked to think of the wave function as a sort of *prediction catalog* [**?**]. This prediction catalogue contains all potential information. Yet, it can be opened only at a *single* particular page. The prediction catalog may be closed before this page is read. Then it could be opened once more at another, complementary, page. By no way it is possible to open the prediction catalog at one page, read and (irreversibly) memorize (measure) the page, and close it; then open it at another, complementary, page. (Two non-complementary pages which correspond to two co-measurable observables can be read simultaneously.)

This may sound a little bit like voodoo. It is tempting to speculate that complementarity can never be modeled by classical metaphors. Yet, classical examples abound. A trivial one is a dark room with a ball moving in it. Suppose that we want to measure its position and its velocity. We first try to measure the ball's position by touching it. This finite contact inevitably causes a finite change of the ball's motion. Therefore, we cannot any longer measure the initial velocity of the ball with arbitrary position.

There are a number of more faithful classical metaphors for quantum complementarity. Take, for instance, Cohen's "firefly-in-a-box" model [?], Wright's urn model [?], as well as Aerts' vessel model [?]. In what follows, we are going to explore a model of complementarity pioneered by Moore [?]. It is based on extremely simple systems—probably the simplest systems you can think of—on finite automata. The finite automata we will consider here are objects which have a finite number of internal states and a finite number of input and output symbols. Their time evolution is mechanistic and can be written down on tables in matrix form. There are no build-in infinities anywhere; no infinite tape or memory, no non-recursive bounds on the runtime *et cetera.*

Let us develop *computational complementarity,* as it is often called [?], as a game between you as the reader and me as the author. The rules of the game are as follows. I first give you all you need to know about the intrinsic workings of the automaton. For example, I tell you, "if the automaton is in state 1 and you input the symbol 2, then the automaton will make a transition into state 2 and output the symbol 0;" and so on. Then I present you a black box which contains a realization of the automaton. The black box has a keyboard, with which you input the input symbols. It has an output display, on which the output symbols appear. No other interfaces are allowed. Suppose that I can choose in which initial state the automaton is at the beginning of the game. I do not tell you this state. Your goal is to find out by experiment which state I have chosen. You can simply guess or relying on your luck by throwing a dice. But

you can also perform clever input-output experiments and analyze your data in order to find out. You win if you give the correct answer. I win if you guess incorrectly. (So, I have to be mean and select worst-case examples).

Suppose that you try very hard. Is cleverness sufficient? Will you always be able to uniquely determine the initial automaton state?

The answer to that question is "no." The reason for this is that there may be situations when the input causes an irreversible transition into a state which does not allow any further queries about the initial state. This is the meaning of the term "self-interference" mentioned above. Any such irreversible loss of information about the initial value of the automaton can be traced back to many-to-one operations [?]: different states are mapped onto a single state with the same output. Many-to-one operations such as "deletion of information" are the only source of entropy increase in mechanistic systems [?, ?].

In the automaton case discussed above, one could, of course, restore reversibility and recover the automaton's initial state by Landauer's "Hänsel und Gretel"-strategy. That is, one could introduce an additional marker at every many-to-one node which indicates the previous state before the transition. But then, as the combined automaton/marker system is reversible, going back to the initial state erases all previous knowledge. This is analogous to the re-opening of pages of Schrödinger's prediction catalog.

Well, this might be a good moment for introducing a sufficiently simple example. Consider, therefore, an automaton which can be in one of three states,

denoted by 1, 2 and 3. This automaton accepts three input symbols, namely 1, 2 and 3. It outputs only two symbols, namely 0 and 1. The transition function of the automaton is as follows: on input 1, it makes a transition to (or remains in) state 1; on input 2, it makes a transition to (or remains in) state 2; on input 3, it makes a transition to (or remains in) state 3. This is a typical irreversible many-to-one operation, since a particular input steers the automaton into that state, no matter in which one of the three possible state it was previously. The output function is also many-to-one and rather simple: whenever both state and input coincide—that is, whenever the guess was correct—it outputs 1; else it outputs 0. So, for example, if it was in state 2 or 3 and receives input 1, it outputs 0 and makes a transition to state 1. There it awaits another input. These automaton specifications can be conveniently represented by diagrams such as the one drawn in Fig. **??**(a).

Computational complementarity manifests itself in the following way: if one does not know the automaton's initial state, one has to make choices between the input of symbols 1,2, or 3; corresponding to definite answers whether the automaton was in state 1, 2 or 3, corresponding to output 1; and (2 or 3), (1 or 3) or (2 or 3), corresponding to output 0, respectively. In the latter case, i.e., whenever the automaton responds with a 0 (for failure), one has lost information about the automaton's initial state, since it surely has made a transition into the state 1,2 or 3. The following propositions can be stated. On input 1, one obtains information that the automaton either was in state 1 (exclusive) or not

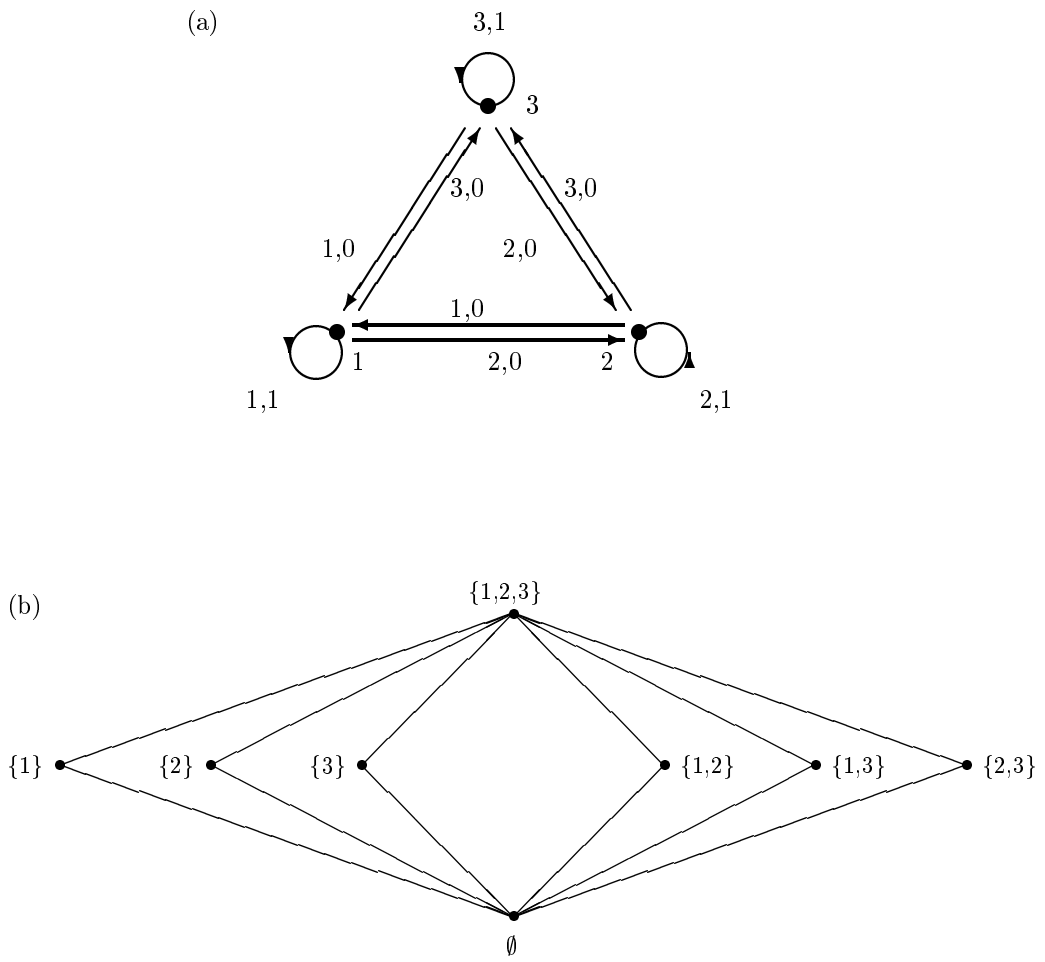

Figure 1: (a) transition diagram of a quantum-like finite automaton featuring computational complementarity. Input and output symbols are separated by a comma. Arrows indicate transitions. (b) Hasse diagram of its propositional structure. Lower elements imply higher ones if they are connected by edge(s).

in state 1, that is, in state 2 or 3. This is denoted by $v(1) = \{\{1\}, \{2,3\}\}$. On input 2, we obtain information that the automaton either was in state 1 (exclusive) or in state 1 or 3, denoted by $v(2) = \{\{2\}, \{1,3\}\}$. On input 3, we obtain information that the automaton either was in state 1 (exclusive) or in state 1 or 2, denoted by $v(3) = \{\{3\}, \{1,2\}\}$. In that way, we naturally arrive at the notion of a *partitioning* of automaton states according to the information obtained from input/output experiments. Every element of the partition stands for the proposition that the automaton is in (one of) the state(s) contained in that partition.

From any partition we can construct the Boolean propositional calculus which is obtained if we identify its atoms with the elements of the partition. We then "paste" all Boolean propositional calculi (sometimes called subalgebras or blocks) together. This is a standard construction in the theory of orthomodular posets [?, ?, ?, ?]. In the above example, we arrive at a form of non-Boolean lattice whose Hasse diagram MO_3 is of the "Chinese latern" type drawn in Fig. ??(b).

Let us go still a little bit further and ask which automaton games of the above kind can people play. This requires the systematic investigation of all possible non-isomorphic automaton propositional structures, or, equivalently, partition logics [?, ?, ?]. In Fig. ??, the Hasse diagrams of all nonisomorphic four-state automaton propositional calculi are drawn.

Recall that the method introduced here is not directly related to diagonal-

Figure 2: Variations of the complementarity game up to four automaton states.

ization and is a second, independent source of undecidability. It is already realizable at an elementary pre-diagonalization level, i.e., without the requirement of computational universality or its arithmetic equivalent. The corresponding machine model is the class of finite automata.

Since any finite state automaton can be simulated by a universal computer, complementarity is a feature of sufficiently complex deterministic universes as well. To put it pointedly: if the physical universe is conceived as the product of a universal computation, then complementarity is an inevitable and necessary feature of the perception of intrinsic observers. It cannot be avoided.

Conversely, any computation can be realized by a sufficiently complex finite automaton. Therefore, the class of all complementary games is a unique one, encompassing all possible deterministic universes.

What has all this to do with intrinsic observers? Well, during the game one is not allowed to "screw open" the black box. Equivalently, one is not allowed to make identical copies of the black box. Both, the "screwing open" operation as well as copying, would represent actions which should only be accessible to "God-like," external observers, but not to intrinsic ones living in their "Cartesian prison." This is similar to quantum mechanics. Copying of quantum information (unless purely classical) or other one-to-many operations are impossible. One cannot, for instance, produce two identical copies of an electron or of a photon [**?**].

The complementarity game described above shows strong similarities to

quantum mechanical systems (in two-dimensional Hilbert space). Indeed, if we could let the black box "shrink" down to point size, we would obtain an analogue of an electron or photon, at least for spin or polarization measurements in three different directions. Suppose we want to measure the spin direction of an electron at some angle φ. We can do this by a Stern-Gerlach device oriented in that particular direction. According to the probabilistic interpretation of the wave function, this measurement "randomizes" (i.e., makes impossible any measurements of the original) spin components in other directions. That is, we loose information about the electron's "original" spin (if it is legitimate to state that it ever had one [?, ?]) along the directions $\varphi' \neq \varphi$. Indeed, the propositional structure of three spin measurements along three different angles is identical to the one drawn in Fig. ??(b). Nevertheless, there is a difference between the "true" quantum particle and its black box-cousin: whereas the former one is supposed to have physical spin or polarization in a *continuity* of directions, the latter one can only be generalized to an arbitrary *countable* number of directions. From a practical point of view, such differences cannot be observed and are therefore operationally irrelevant [?, ?].

Even to high-ranking specialists, quantum mechanical effects appear mind-boggling [?]. Amazingly enough, the complementarity game based on automata beats quantum mechanics by weirder peculiarities. Take, as an example, the complementarity game with the automaton drawn in Fig. ??(a). Input of the sequence of two symbols 00 decides between the automaton states 1 and 2 and 3

or 4. The resulting partition is $v(00) = \{\{1\}, \{2\}, \{3, 4\}\}$. Input of the sequence of two symbols 10 decides between the automaton states 1 or 2 and 3 and 4. The resulting partition is $v(10) = \{\{1, 2\}, \{3\}, \{4\}\}$. By pasting these two blocks together, we obtain a propositional structure represented in Fig. **??**(b).

This complementarity game has several peculiar features. It is no lattice because the supremum and infimum is not uniquely definable. The "implication" is not transitive either, because $1 \to 1 \vee 2$ requires input 00 and $1 \vee 2 \to 1 \vee 2 \vee 3$ requires input 10, whereas $1 \to 1 \vee 2 \vee 3$ cannot be realized by any experiment.

It would be nice if some day the experimenters would find physical systems which behave in that way. Then, of course, we would have to abandon quantum mechanics and learn some theory of the complementarity game.

4 Should physicists really bother with undecidability?

The program to study relative limits of knowledge can be attacked from two opposite and extreme positions. On the one hand, it may be objected that there are no principally unknowables, because everything is strictly causal. On the other hand it may be stated that undecidability in physics is a trivial matter of fact and must be accepted without any further efforts.

The first, rationalistic, position is based on the Cartesian assumption that the world is totally and in all of its aspects conceivable and predictable by (hu-

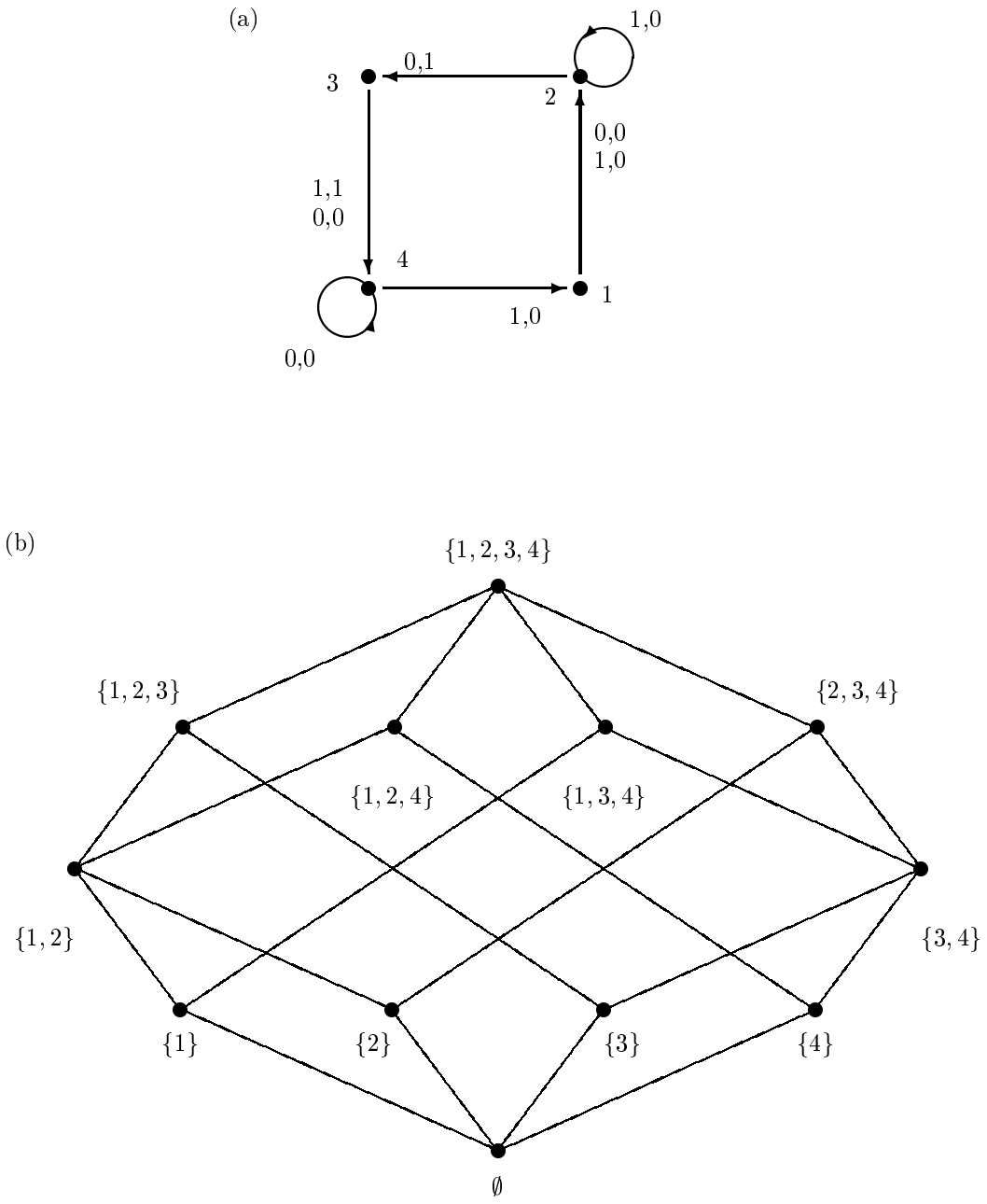

Figure 3: (a) complementarity game featuring weirder-than quantum properties;

(b) Hasse diagram of its propositional structure.

man) rational thought. Laplace's demon [?] is a metaphor for this attitude. Indeed, to many physicists, undecidability and unpredictability are everyday phenomena. They encounter a problem which they cannot solve or ask questions they cannot answer. Yet, they would hardly conceive this experience as an indication that there is something out there which is profoundly undecidable. It might not be unfair to state that—with the remarkable exceptions of chaos and quantum theory—most physicists perceive undecidable statements not as fundamentally unknowable but as a challenge and a possibility for future knowledge.

Quite similarly, some mathematicians tend to perceive Gödel's incompleteness theorems as artifacts. To them, Gödelian sentences appear curious, even dubious, and explicitly constructed for their purpose. Despite proofs that "almost all" true theorems are undecidable [?], they feel that all "real" mathematical problems they bother with *are* solvable.

In this century, the second, irrational approach, has most influentially been publicized—despite many reluctances from leading quantum pioneers [?, ?]—by the Copenhagen interpretation of quantum mechanics. Chaos theory, as it is often called, has given irrationality a further kick. Already in 1889, Poincaré suggested that certain n-body problems may turn out to be impossible to solve [?]. Even today, after the development of recursive (computable) function theory, many issues remain unsettled. Certain assumptions and problems yield the nonpreservation of computability in classical analysis [?, ?, ?, ?, ?, ?]. The ne-

cessity and the physical relevance of the classical continuum is at least debatable [?].

We propose here to pursue a third path. This third path is characterized by a formal investigation of the descriptive limits of theories, as well as of predictability in general. It may well be that this is a further, necessary step we have to go in our rational understanding of nature.

At the end, let me summarize and come back to the question posed before, "undecidability everywere?" It may well be that yes, there is undecidability everywhere, and that we are confronted with it very often. We just may not have identified undecidability correctly, as some emerging feature of (self-) description (with-) in a mechanistic universe. Depending on our philosophical assumptions, some of us may like to think that the everday unknowns are either manifestations of some basic randomness, a sort of "chaos" underlying nature; or, on the contrary, that they are simply an artifact of our limited knowledge and power to implement that knowledge. We may also realize that there is yet another possibility, having to do with with the fact that, informally stated, self-knowledge is necessarily incomplete.

References

[1] In a double slit experiment, the wave function Ψ is actually in a coherent superposition $\Psi = \psi(1) + \psi(2)$ of the amplitude for the particle to go

through slit 1 and slit 2.

[2] D. Aerts, *Example of a macroscopic classical situation that violates Bell inequalities.* Lettere al Nuovo Cimento **34** (1982), 107–111.

[3] C. H. Bennett, *Logical Reversibility of Computation*, *IBM J. Res. Dev.* **17**, 525-532 (1973); reprinted in: *Maxwell's Demon*, ed. by H. S. Leff and A. F. Rex (Princeton University Press, 1990), pp. 197-204.

[4] J. Bernstein, *Quantum Profiles* (Princeton University Press, Princeton, 1991), pp. 140-141.

[5] The *Bible* contains a passage, which refers to Epimenides, a Crete living in the capital city of Cnossus: *"One of themselves, a prophet of their own, said, 'Cretans are always liars, evil beasts, lazy gluttons.' "* — St. Paul, Epistle to Titus I (12-13). For more details, see A. R. Anderson, *St. Paul's epistle to Titus*, in *The Paradox of the Liar*, ed. by R. L. Martin (Yale University Press, New Haven, 1970).

[6] R. J. Boskovich, *De spacio et tempore, ut a nobis cognoscuntur* (Vienna, 1755); English translation in *A Theory of Natural Philosophy*, ed. by J. M. Child (Open Court, Chicago, 1922; reprinted by MIT press, Cambridge, MA, 1966), p. 203-205.

[7] A. H. Brady, *The Busy Beaver Game and the Meaning of Life*, in *The Universal Turing Machine. A Half-Century Survey*, ed. by R. Herken (Kammerer & Unverzagt, Hamburg, 1988), p. 259.

[8] P. W. Bridgman, *A Physicists Second Reaction to Mengenlehre, Scripta Mathematica* **2**, 101-117; 224-234 (1934); cf. R. Landauer, *Advertisement For a Paper I Like*, in *On Limits*, ed. by J. L. Casti and J. F. Traub (Santa Fe Institute Report 94-10-056, Santa Fe, NM, 1994), p.39; P. W. Bridgman, *Reflections of a Physicist* (Philosophical Library, New York, 1950).

[9] C. Calude, *Information and Randomness — An Algorithmic Perspective* (Springer, Berlin, 1994).

[10] C. Calude, D. I. Campbell, K. Svozil and D. Ştefănescu, *Strong Determinism vs. Computability*, *e-print quant-ph/9412004*.

[11] C. Calude, H. Jürgensen and M. Zimand, *Is independence an exception? Appl. Math. Comput.* **66**, 63-76 (1994).

[12] J. L. Casti, *Beyond Believe: Randomness, Prediction and Explanation in Science* (CRC Press, Boca Raton, Florida, 1990); *Searching for Certainty: What Scientists Can Know about the Future* (W. Morrow& Company, New York, 1992).

[13] G. J. Chaitin, *Information, Randomness and Incompleteness, Second edition* (World Scientific, Singapore, 1987, 1990); *Algorithmic Information Theory* (Cambridge University Press, Cambridge, 1987); *Information-Theoretic Incompleteness* (World Scientific, Singapore, 1992).

[14] G. J. Chaitin, *Journal of the Assoc. Comput. Mach.* **21**, 403 (1974); reprinted in [**?**].

[15] G. J. Chaitin, *Computing the Busy Beaver Function*, in *Open Problems in Communication and Computation*, ed. by T. M. Cover and B. Gopinath (Springer, New York, 1987), p. 108; reprinted in [**?**].

[16] D.W. Cohen, *An Introduction to Hilbert Space and Quantum Logic.* Springer–Verlag, New York, Berlin, Heidelberg, London, Paris, Tokyo, 1989.

[17] N. C. A. da Costa and F. A. Doria, *Foundations of Physics Letters* **4**, 363 (1991); N. C. A. da Costa and F. A. Doria, *International Journal of Theoretical Physics* **30**, 1041 (1991).

[18] M. Davis, *Computability & Unsolvability* (McGraw-Hill, New York, 1958).

[19] A. K. Dewdney, *Scientific American* **251**, 10 (July 1984).

[20] D. Finkelstein, S. R. Finkelstein, *Computational complementarity.* Inter. J. Theor. Phys. **22** (1983), 753–779.

[21] Ph. Frank, *Das Kausalgesetz und seine Grenzen* (Springer, Vienna 1932).

[22] K. Gödel, *Monatshefte für Mathematik und Physik* **38**, 173 (1931); English translation in [**?**] and in Davis, ref. [**?**].

[23] K. Gödel, *Collected Works, Volume I, Publications 1929-1936*, ed. by S. Feferman, J. W. Dawson, Jr., St. C. Kleene, G. H. Moore, R. M. Solovay, J. van Heijenoort (Oxford University Press, Oxford, 1986).

[24] E. M. Gold, *Information and Control* **10**, 447 (1967).

[25] D. B. Greenberger, M. Horne and A. Zeilinger *Physics Today* **46**, 22 (August 1993).

[26] N. Herbert, *Foundation of Physics* **12**, 1171 (1982); W. K. Wooters and W. H. Zurek, *Nature* **299**, 802 (1982); P. W. Milonni and M. L. Hardies, *Phys. Lett.* **92A**, 321 (1982); R. J. Glauber, *Amplifiers, Attenuators and the Quantum Theory of Measurement*, in *Frontiers in Quantum Optics*, ed. by E. R. Pikes and S. Sarkar (Adam Hilger, Bristol 1986) C. M. Caves, *Phys. Rev.* **D 26**, 1817 (1982).

[27] D. Hilbert, *Über das Unendliche, Math. Annalen* **95**, 161-190 (1926).

[28] M. Jammer, *The Philosophy of Quantum Mechanics* (John Wiley, New York, 1974).

[29] G. Kalmbach, *Orthomodular Lattices* (Academic Press, New York, 1983); *Measures and Hilbert Lattices* (World Scientific, Singapore, 1986); *Foundations of Physics* **20**, 801 (1990).

[30] G. Kreisel, A notion of mechanistic theory, *Synthese* 29(1974), 11-26.

[31] R. Landauer, *Irreversibility and Heat Generation in the Computing Process*, *IBM J. Res. Dev.* **5**, 183-191 (1961); reprinted in: *Maxwell's Demon*, ed. by H. S. Leff and A. F. Rex (Princeton University Press, 1990), pp. 188-196; *Fundamental Physical Limitations of the Computational Process; an Informal Commentary*, in *Cybernetics Machine Group Newsheet* 1/1/87;

Computation, Measurement, Communication and Energy Dissipation, in *Selected Topics in Signal Processing*, ed. by S. Haykin (prentice Hall, Englewood Cliffs, NJ, 1989), p. 18; *Physics Today* **44**, 23 (Mai 1991); *Zig-Zag Path to Understanding*, *Proceedings of the Workshop on Physics and Computation PHYSCOMP '94* (IEEE Press, Los Alamitos, CA, 1994), pp. 54-59.

[32] P. S. Laplace, *Théorie analytique des probabiletés. Introduction*, in *Oeuvres de Laplace* **7**, 6 (1847).

[33] N. D. Mermin, *Rev. Mod. Phys.* **65**, 803 (1993).

[34] Ch. D. Moore, *Phys. Rev. Lett.* **64**, 2354 (1990).

[35] E. F. Moore, *Gedanken–experiments on sequential machines.* In: Automata Studies, eds. C.E. Shannon, J.McCarthy, Princeton Univ. Press, Princeton, N.J. 1956.

[36] M. Navara and V. Rogalewicz, *Math. Nachr.* **154**, 157 (1991).

[37] P. Odifreddi, *Classical Recursion Theory* (North-Holland, Amsterdam, 1989).

[38] A. Peres, *Am. J. Phys.* **46**, 745 (1978).

[39] A. Peres, *Quantum Theory: Concepts & Methods* (Kluwer Academic Publishers, Dordrecht, 1993).

[40] R. Piziak, *Orthomodular lattices and quadratic spaces: a survey*, *Rocky Mountain Journal of Mathematics* **21**, 951 (1991).

[41] H. Poincaré, *Sur le problème des trois corps*, *Acta Mathematica* **13**, 5-271 (1890).

[42] K. R. Popper, *The British Journal for the Philosophy of Science* **1**, 117, 173 (1950).

[43] M. Pour-El, I. Richards, *Computability in Analysis and Physics*, (Springer-Verlag, Berlin, 1989). For a critique, see D. S. Bridges, *Constructive mathematics and unbounded operators—a reply to Hellman, J. Phil. Logic*, in press.

[44] P. Pták and S. Pulmannová, *Orthomodular Structures as Quantum Logics* (Kluwer Academic Publishers, Dordrecht, 1991).

[45] H. Putnam, *Reason, Truth and History* (Cambridge University Press, Cambridge, 1981).

[46] T. Rado, *Bell Sys. T.* (May 1962), 877.

[47] M. Rasetti, *Chaos, Solitons & Fractals* **5**, 133-138 (1995).

[48] H. Rogers, *Theory of Recursive Functions and Effective Computability* (MacGraw-Hill, New York 1967).

[49] O. E. Rössler, *Endophysics*, in *Real Brains, Artificial Minds*, ed. by J. L. Casti and A. Karlquist (North-Holland, New York, 1987), p. 25.

[50] O. E. Rössler, *Endophysics, Die Welt des inneren Beobachters*, ed. by P. Weibel (Merwe Verlag, Berlin, 1992).

[51] R. Rucker, *Infinity and the Mind* (Birkhäuser, Boston 1982; reprinted by Bantam Books 1986).

[52] M. Schaller, K. Svozil, *Partition logics of automata.* Il Nuovo Cimento **109 B** (1994), 167–176; *Automaton partition logic versus quantum logic,* Inter. J. Theor. Phys, **34**, 1741-1750 (1995). ; *Automaton logic,* Inter. J. Theor. Phys, *in print.*

[53] E. Schrödinger, *Naturwissenschaften* **23**, 807; 823; 844 (1935) [English translation in [**?**] p. 152].

[54] E. P. Specker, *Dialectica* **14**, 175 (1960); S. Kochen and E. P. Specker, *The calculus of partial propositional functions,* in *Proceedings of the 1964 International Congress for Logic, Methodology and Philosophy of Science, Jerusalem* (North Holland, Amsterdam, 1965), p. 45-57; S. Kochen and E. P. Specker, *Logical Structures arising in quantum theory*, in *Symposium on the Theory of Models, Proceedings of the 1963 International Symposium at Berkeley* (North Holland, Amsterdam, 1965), p. 177-189; S. Kochen and E. P. Specker, *Journal of Mathematics and Mechanics* **17**, 59 (1967); reprinted in E. Specker, *Selecta* (Birkhäuser Verlag, Basel, 1990); Erna Clavadetscher-Seeberger, *Eine partielle Prädikatenlogik* (Dissertation, ETH-Zürich, Zürich, 1983); N. D. Mermin [**?**].

[55] D. Ştefănescu, *Mathematical Models in Physics*, University of Bucharest Press, 1984. (in Romanian)

[56] K. Svozil, *On the setting of scales for space and time in arbitrary quantized media* (Lawrence Berkeley Laboratory preprint LBL-16097, May 1983).

[57] K. Svozil, *Europhysics Letters* **2**, 83 (1986).

[58] K. Svozil, *Il Nuovo Cimento* **96B**, 127 (1986).

[59] K. Svozil, *Randomness & Undecidability in Physics* (World Scientific, Singapore, 1993).

[60] K. Svozil, "Extrinsic-Intrinsic Concept and Complementarity", in *Inside Versus Outside*, ed. by H. Atmanspacher, G. J. Dalenoort (Springer, Berlin, 1994), p. 273-288.

[61] K. Svozil, *Set Theory and Physics*, *Foundations of Physics*, *in print.*

[62] K. Svozil and R. R. Zapatrin, *Empirical logic of finite automata: microstatements versus macrostatements*, *submitted.*

[63] A. Tarski, *Der Wahrheitsbegriff in den Sprachen der deduktiven Disziplinen*, in *Akademie der Wissenschaften in Wien, Mathematisch-naturwissenschaftliche Klasse, Anzeiger* **69**, 24 (1932).

[64] T. Toffoli, *The role of the observer in uniform systems*, in *Applied General Systems Research*, ed. by G. Klir (Plenum Press, New York, London, 1978).

[65] A. M. Turing, *Proc. London Math. Soc.* **(2), 42**, 230 (1936-7), reprinted in [**?**].

[66] J. von Neumann, *Theory of Self-Reproducing Automata*, ed. by A. W. Burks (University of Illinois Press, Urbana, 1966).

[67] P. S. Wang, The undecidability of the existence of zeros of real elementary functions, *J. Assoc. Comput. Mach.* **21** (1974), 586-589.

[68] J. A. Wheeler and W. H. Zurek, eds.., *Quantum Theory and Measurement* (Princeton University Press, Princeton, 1983).

[69] St. Wolfram, *Cellular Automata and Complexity, Collected Papers* (Addison-Wesley, Reading, MA, 1994).

[70] R. Wright, *Generalized urn models.* Found. Phys. **20** (1990), 881–903.

Contents