Gödel Machines: Fully Self-referential Optimal Universal Self-improvers*

Jürgen Schmidhuber

IDSIA, Galleria 2, 6928 Manno (Lugano), Switzerland &
TU Munich, Boltzmannstr. 3, 85748 Garching, München, Germany
juergen@idsia.ch - http://www.idsia.ch/~juergen

Summary. We present the first class of mathematically rigorous, general, fully self-referential, self-improving, optimally efficient problem solvers. Inspired by Kurt Gödel's celebrated self-referential formulas (1931), such a problem solver rewrites any part of its own code as soon as it has found a proof that the rewrite is *useful,* where the problem-dependent *utility function* and the hardware and the entire initial code are described by axioms encoded in an initial proof searcher which is also part of the initial code. The searcher systematically and efficiently tests computable *proof techniques* (programs whose outputs are proofs) until it finds a provably useful, computable self-rewrite. We show that such a self-rewrite is globally optimal—no local maxima!—since the code first had to prove that it is not useful to continue the proof search for alternative self-rewrites. Unlike previous *non*-self-referential methods based on hardwired proof searchers, ours not only boasts an optimal *order* of complexity but can optimally reduce any slowdowns hidden by the $O()$-notation, provided the utility of such speed-ups is provable at all.

1 Introduction and Outline

In 1931 Kurt Gödel used elementary arithmetics to build a universal programming language for encoding arbitrary proofs, given an arbitrary enumerable set of axioms. He went on to construct *self-referential* formal statements that claim their own unprovability, using Cantor's diagonalization trick [5] to demonstrate that formal systems such as traditional mathematics are either flawed in a certain sense or contain unprovable but true statements [11]. Since Gödel's exhibition of the fundamental limits of proof and computation, and Konrad Zuse's subsequent construction of the first working programmable computer (1935-1941), there has been a lot of work on specialized algorithms solving problems taken from more or less general problem classes. Apparently, however, one remarkable fact has so far escaped the attention of computer scientists: it is possible to use self-referential proof systems to build optimally efficient yet conceptually very simple universal problem solvers.

All traditional algorithms for problem solving / machine learning / reinforcement learning [19] are hardwired. Some are designed to improve some limited type of policy through experience, but are not part of the modifiable

*Certain parts of this work appear in [46] and [47], both by Springer.

policy, and cannot improve themselves in a theoretically sound way. Humans are needed to create new/better problem solving algorithms and to prove their usefulness under appropriate assumptions.

Let us eliminate the restrictive need for human effort in the most general way possible, leaving all the work including the proof search to a system that can rewrite and improve itself in arbitrary computable ways and in a most efficient fashion. To attack this *"Grand Problem of Artificial Intelligence,"* we introduce a novel class of optimal, fully self-referential [11] general problem solvers called *Gödel machines* [43].[1] They are universal problem solving systems that interact with some (partially observable) environment and can in principle modify themselves without essential limits besides the limits of computability. Their initial algorithm is not hardwired; it can completely rewrite itself, but only if a proof searcher embedded within the initial algorithm can first prove that the rewrite is useful, given a formalized utility function reflecting computation time and expected future success (e.g., rewards). We will see that self-rewrites due to this approach are actually *globally optimal* (Theorem 1, Section 4), relative to Gödel's well-known fundamental restrictions of provability [11]. These restrictions should not worry us; if there is no proof of some self-rewrite's utility, then humans cannot do much either.

The initial proof searcher is $O()$-optimal (has an optimal order of complexity) in the sense of Theorem 2, Section 5. Unlike Hutter's hardwired systems [17, 16] (Section 2), however, a Gödel machine can further speed up its proof searcher to meet *arbitrary* formalizable notions of optimality beyond those expressible in the $O()$-notation. Our approach yields the first theoretically sound, fully self-referential, optimal, general problem solvers.

Outline. Section 2 presents basic concepts, relations to the most relevant previous work, and limitations. Section 3 presents the essential details of a self-referential axiomatic system, Section 4 the Global Optimality Theorem 1, and Section 5 the $O()$-optimal (Theorem 2) initial proof searcher. Section 6 provides examples and additional relations to previous work, briefly discusses issues such as a *technical* justification of consciousness, and provides answers to several frequently asked questions about Gödel machines.

2 Basic Overview, Relation to Previous Work, and Limitations

Many traditional problems of computer science require just one problem-defining input at the beginning of the problem solving process. For example, the initial input may be a large integer, and the goal may be to factorize it. In what follows, however, we will also consider the *more general case* where

[1] Or *'Goedel machine'*, to avoid the *Umlaut*. But *'Godel machine'* would not be quite correct. Not to be confused with what Penrose calls, in a different context, *'Gödel's putative theorem-proving machine'* [29]!

the problem solution requires interaction with a dynamic, initially unknown environment that produces a continual stream of inputs and feedback signals, such as in autonomous robot control tasks, where the goal may be to maximize expected cumulative future reward [19]. This may require the solution of essentially arbitrary problems (examples in Sect. 6.2 formulate traditional problems as special cases).

2.1 Notation and Set-up

Unless stated otherwise or obvious, throughout the paper newly introduced variables and functions are assumed to cover the range implicit in the context. B denotes the binary alphabet $\{0,1\}$, B^* the set of possible bitstrings over B, $l(q)$ denotes the number of bits in a bitstring q; q_n the n-th bit of q; λ the empty string (where $l(\lambda)=0$); $q_{m:n}=\lambda$ if $m>n$ and $q_m q_{m+1}\ldots q_n$ otherwise (where $q_0 := q_{0:0} := \lambda$).

Our hardware (e.g., a universal or space-bounded Turing machine or the abstract model of a personal computer) has a single life which consists of discrete cycles or time steps $t=1,2,\ldots$. Its total lifetime T may or may not be known in advance. In what follows, the value of any time-varying variable Q at time t will be denoted by $Q(t)$. Occasionally it may be convenient to consult Fig. 1.

During each cycle our hardware executes an elementary operation which affects its variable state $s \in \mathcal{S} \subset \mathcal{B}^*$ and possibly also the variable environmental state $Env \in \mathcal{E}$. (Here we need not yet specify the problem-dependent set $\mathcal{E}$). There is a hardwired state transition function $F : \mathcal{S}\times\mathcal{E} \to \mathcal{S}$. For $t>1$, $s(t) = F(s(t-1), Env(t-1))$ is the state at a point where the hardware operation of cycle $t-1$ is finished, but the one of t has not started yet. $Env(t)$ may depend on past output actions encoded in $s(t-1)$ and is simultaneously updated or (probabilistically) computed by the possibly reactive environment.

In order to conveniently talk about programs and data, we will often attach names to certain string variables encoded as components or substrings of s. Of particular interest are 3 variables called *time*, x, y, p:

1. At time t, variable *time* holds a unique binary representation of t. We initialize $time(1)$ = '1', the bitstring consisting only of a one. The hardware increments *time* from one cycle to the next. This requires at most $O(log\ t)$ and on average only $O(1)$ computational steps.
2. Variable x holds environmental inputs. For $t>1$, $x(t)$ may differ from $x(t-1)$ only if a program running on the Gödel machine has executed a special input-requesting instruction at time $t-1$. Generally speaking, the delays between successive inputs should be sufficiently large so that programs can perform certain elementary computations on an input, such as copying it into internal storage (a reserved part of s) before the next input arrives.

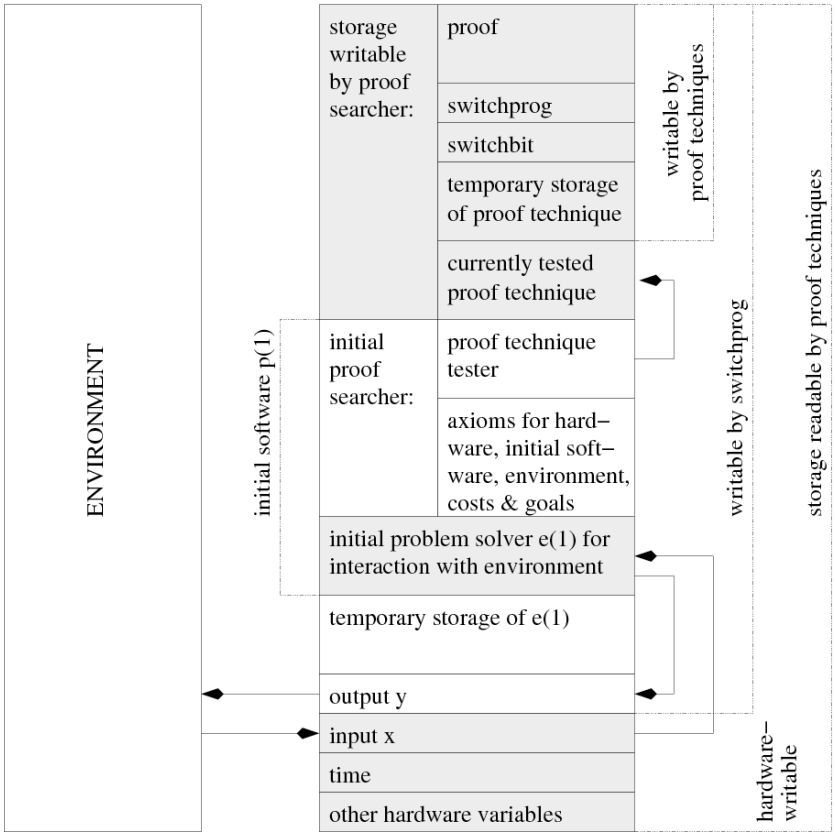

Fig. 1: Storage snapshot of a not yet self-improved example Gödel machine, with the initial software still intact. See text for details.

3. $y(t)$ is an output bitstring which may subsequently influence the environment, where $y(1) =$ '0' by default. For example, $y(t)$ could be interpreted as a control signal for an environment-manipulating robot whose actions may have an effect on future inputs.
4. $p(1)$ is the initial software: a program implementing the original policy for interacting with the environment and for proof searching. Details will be discussed below.

At any given time t ($1 \leq t \leq T$) the goal is to maximize future success or *utility*. A typical *"value to go"* utility function (to be maximized) is of the form $u(s, Env) : \mathcal{S} \times \mathcal{E} \rightarrow \mathcal{R}$, where $\mathcal{R}$ is the set of real numbers:

$$u(s, Env) = E_\mu \left[\sum_{\tau=time}^{T} r(\tau) \;\middle|\; s, Env \right], \tag{1}$$

where $r(t)$ is a real-valued reward input (encoded within $s(t)$) at time t, $E_\mu(\cdot \mid \cdot)$ denotes the conditional expectation operator with respect to some possibly unknown distribution μ from a set M of possible distributions (M reflects whatever is known about the possibly probabilistic reactions of the environment), and the above-mentioned $time = time(s)$ is a function of state s which uniquely identifies the current cycle. Note that we take into account the possibility of extending the expected lifespan $E_\mu(T \mid s, Env)$ through appropriate actions.

Alternative formalizable utility functions could favor improvement of *worst case* instead of *expected* future performance, or higher reward intake *per time interval* etc.

2.2 Basic Idea of Gödel Machine

Our machine becomes a self-referential [11] *Gödel machine* by loading it with a particular form of machine-dependent, self-modifying code p. The initial code $p(1)$ at time step 1 includes a (typically sub-optimal) problem solving subroutine for interacting with the environment, such as Q-learning [19], and a general proof searcher subroutine (Section 5) that systematically makes pairs *(switchprog, proof)* (variable substrings of s) until it finds a *proof* of a target theorem which essentially states: *'the immediate rewrite of* p *through current program* switchprog *on the given machine implies higher utility than leaving* p *as is'.* Then it executes *switchprog*, which may completely rewrite p, including the proof searcher. Section 3 will explain details of the necessary initial axiomatic system $\mathcal{A}$ encoded in $p(1)$.
The Global Optimality Theorem (Theorem 1, Sect. 4) shows this self-improvement strategy is not greedy: since the utility of *'leaving p as is'* implicitly evaluates all possible alternative *switchprog*s which an unmodified p might find later, we obtain a globally optimal self-change—the *current switchprog* represents the best of all possible relevant self-changes, relative to the given resource limitations and initial proof search strategy.

2.3 Proof Techniques and an $O()$-optimal Initial Proof Searcher.

Section 5 will present an $O()$-optimal initialization of the proof searcher, that is, one with an optimal *order* of complexity (Theorem 2). Still, there will remain a lot of room for self-improvement hidden by the $O()$-notation. The searcher uses an online extension of *Universal Search* [23, 25] to systematically test *online proof techniques*, which are proof-generating programs that may read parts of state s (similarly, mathematicians are often more interested in proof techniques than in theorems). To prove target theorems as above, proof

techniques may invoke special instructions for generating axioms and applying inference rules to prolong the current *proof* by theorems. Here an axiomatic system $\mathcal{A}$ encoded in $p(1)$ includes axioms describing *(a)* how any instruction invoked by a program running on the given hardware will change the machine's state s (including instruction pointers etc.) from one step to the next (such that proof techniques can reason about the effects of any program including the proof searcher), *(b)* the initial program $p(1)$ itself (Section 3 will show that this is possible without introducing circularity), **(c)** stochastic environmental properties, **(d)** the formal utility function u, e.g., equation (1). The evaluation of utility automatically takes into account computational costs of all actions including proof search.

2.4 Relation to Hutter's Previous Work

Hutter's non-self-referential but still $O()$-optimal *'fastest' algorithm for all well-defined problems* HSEARCH [17] uses a *hardwired* brute force proof searcher. Assume discrete input/output domains X/Y, a formal problem specification $f : X \to Y$ (say, a functional description of how integers are decomposed into their prime factors), and a particular $x \in X$ (say, an integer to be factorized). HSEARCH orders all proofs of an appropriate axiomatic system by size to find programs q that for all $z \in X$ provably compute $f(z)$ within time bound $t_q(z)$. Simultaneously it spends most of its time on executing the q with the best currently proven time bound $t_q(x)$. It turns out that HSEARCH is as fast as the *fastest* algorithm that provably computes $f(z)$ for all $z \in X$, save for a constant factor smaller than $1 + \epsilon$ (arbitrary $\epsilon > 0$) and an f-specific but x-independent additive constant [17]. This constant may be enormous though.

Hutter's AIXI*(t,l)* [16] is related. In discrete cycle $k = 1, 2, 3, \ldots$ of AIXI*(t,l)*'s lifetime, action $y(k)$ results in perception $x(k)$ and reward $r(k)$, where all quantities may depend on the complete history. Using a universal computer such as a Turing machine, AIXI*(t,l)* needs an initial offline setup phase (prior to interaction with the environment) where it uses a *hardwired* brute force proof searcher to examine all proofs of length at most L, filtering out those that identify programs (of maximal size l and maximal runtime t per cycle) which not only could interact with the environment but which for all possible interaction histories also correctly predict a lower bound of their own expected future reward. In cycle k, AIXI*(t,l)* then runs all programs identified in the setup phase (at most 2^l), finds the one with highest self-rating, and executes its corresponding action. The problem-independent setup time (where almost all of the work is done) is $O(L \cdot 2^L)$. The online time per cycle is $O(t \cdot 2^l)$. Both are constant but typically huge.

Advantages and Novelty of the Gödel Machine. There are major differences between the Gödel machine and Hutter's HSEARCH [17] and AIXI*(t,l)* [16], including:

1. The theorem provers of HSEARCH and AIXI*(t,l)* are hardwired, non-self-referential, unmodifiable meta-algorithms that cannot improve them-

selves. That is, they will always suffer from the same huge constant slowdowns (typically $\gg 10^{1000}$) buried in the $O()$-notation. But there is nothing in principle that prevents our truly self-referential code from proving and exploiting drastic reductions of such constants, in the best possible way that provably constitutes an improvement, if there is any.

2. The demonstration of the $O()$-optimality of HSEARCH and AIXI*(t,l)* depends on a clever allocation of computation time to some of their unmodifiable meta-algorithms. Our Global Optimality Theorem (Theorem 1, Section 4), however, is justified through a quite different type of reasoning which indeed exploits and crucially depends on the fact that there is no unmodifiable software at all, and that the proof searcher itself is readable and modifiable and can be improved. This is also the reason why its self-improvements can be more than merely $O()$-optimal.
3. HSEARCH uses a "trick" of proving more than is necessary which also disappears in the sometimes quite misleading $O()$-notation: it wastes time on finding programs that provably compute $f(z)$ for all $z \in X$ even when the current $f(x)(x \in X)$ is the only object of interest. A Gödel machine, however, needs to prove only what is relevant to its goal formalized by u. For example, the general u of eq. (1) completely ignores the limited concept of $O()$-optimality, but instead formalizes a stronger type of optimality that does not ignore huge constants just because they are constant.
4. Both the Gödel machine and AIXI*(t,l)* can maximize expected reward (HSEARCH cannot). But the Gödel machine is more flexible as we may plug in *any* type of formalizable utility function (e.g., *worst case* reward), and unlike AIXI*(t,l)* it does not require an enumerable environmental distribution.

Nevertheless, we may use AIXI*(t,l)* or HSEARCH to initialize the substring e of p which is responsible for interaction with the environment. The Gödel machine will replace e as soon as it finds a provably better strategy.

2.5 Limitations of Gödel Machines

The fundamental limitations are closely related to those first identified by Gödel's celebrated paper on self-referential formulae [11]. Any formal system that encompasses arithmetics (or ZFC, etc.) is either flawed or allows for unprovable but true statements. Hence, even a Gödel machine with unlimited computational resources must ignore those self-improvements whose effectiveness it cannot prove, e.g., for lack of sufficiently powerful axioms in $\mathcal{A}$. In particular, one can construct pathological examples of environments and utility functions that make it impossible for the machine to ever prove a target theorem. Compare Blum's speed-up theorem [3, 4] based on certain incomputable predicates. Similarly, a realistic Gödel machine with limited resources cannot profit from self-improvements whose usefulness it cannot prove within its time and space constraints.

Nevertheless, unlike previous methods, it can in principle exploit at least the *provably* good speed-ups of *any* part of its initial software, including those parts responsible for huge (but problem class-independent) slowdowns ignored by the earlier approaches [17, 16].

3 Essential Details of One Representative Gödel Machine

Theorem proving requires an axiom scheme yielding an enumerable set of axioms of a formal logic system $\mathcal{A}$ whose formulas and theorems are symbol strings over some finite alphabet that may include traditional symbols of logic (such as $\rightarrow, \wedge, =, (,), \forall, \exists, \ldots$, $c_1, c_2, \ldots$, $f_1, f_2, \ldots$), probability theory (such as $E(\cdot)$, the expectation operator), arithmetics $(+, -, /, =, \sum, <, \ldots)$, string manipulation (in particular, symbols for representing any part of state s at any time, such as $s_{7:88}(5555)$). A proof is a sequence of theorems, each either an axiom or inferred from previous theorems by applying one of the inference rules such as *modus ponens* combined with *unification*, e.g., [10].

The remainder of this chapter will omit standard knowledge to be found in any proof theory textbook. Instead of listing *all* axioms of a particular $\mathcal{A}$ in a tedious fashion, we will focus on the novel and critical details: how to overcome problems with self-reference and how to deal with the potentially delicate online generation of proofs that talk about and affect the currently running proof generator itself.

3.1 Proof Techniques

Brute force proof searchers (used in Hutter's AIXI*(t,l)* and HSEARCH; see Section 2.4) systematically generate all proofs in order of their sizes. To produce a certain proof, this takes time exponential in proof size. Instead our $O()$-optimal $p(1)$ will produce many proofs with low algorithmic complexity [52, 21, 26] much more quickly. It systematically tests (see Sect. 5) *proof techniques* written in universal language $\mathcal{L}$ implemented within $p(1)$. For example, $\mathcal{L}$ may be a variant of PROLOG [7] or the universal FORTH[28]-inspired programming language used in recent work on optimal search [45]. A proof technique is composed of instructions that allow any part of s to be read, such as inputs encoded in variable x (a substring of s) or the code of $p(1)$. It may write on s^p, a part of s reserved for temporary results. It also may rewrite *switchprog*, and produce an incrementally growing proof placed in the string variable *proof* stored somewhere in s. *proof* and s^p are reset to the empty string at the beginning of each new proof technique test. Apart from standard arithmetic and function-defining instructions [45] that modify s^p, the programming language $\mathcal{L}$ includes special instructions for prolonging the current *proof* by correct theorems, for setting *switchprog*, and for checking whether a provably optimal p-modifying program was found and should be executed now. Certain long proofs can be produced by short proof techniques.

The nature of the six *proof*-modifying instructions below (there are no others) makes it impossible to insert an incorrect theorem into *proof*, thus trivializing proof verification:

1. **get-axiom(n)** takes as argument an integer n computed by a prefix of the currently tested proof technique with the help of arithmetic instructions such as those used in previous work [45]. Then it appends the n-th axiom (if it exists, according to the axiom scheme below) as a theorem to the current theorem sequence in *proof*. The initial axiom scheme encodes:
   a) **Hardware axioms** describing the hardware, formally specifying how certain components of s (other than environmental inputs x) may change from one cycle to the next.
   For example, if the hardware is a Turing machine[2] (TM) [56], then $s(t)$ is a bitstring that encodes the current contents of all tapes of the TM, the positions of its scanning heads, and the current *internal state* of the TM's finite state automaton, while F specifies the TM's look-up table which maps any possible combination of internal state and bits above scanning heads to a new internal state and an action such as: replace some head's current bit by 1/0, increment (right shift) or decrement (left shift) some scanning head, read and copy next input bit to cell above input tape's scanning head, etc. Alternatively, if the hardware is given by the abstract model of a modern microprocessor with limited storage, $s(t)$ will encode the current storage contents, register values, instruction pointers, etc.
   For example, the following axiom could describe how some 64-bit hardware's instruction pointer stored in $s_{1:64}$ is continually incremented as long as there is no overflow and the value of s_{65} does not indicate that a jump to some other address should take place:

$$(\forall t \forall n : [(n < 2^{64} - 1) \wedge (n > 0) \wedge (t > 1) \wedge (t < T)$$

$$\wedge (string2num(s_{1:64}(t)) = n) \wedge (s_{65}(t) = \text{'0'})]$$

$$\rightarrow (string2num(s_{1:64}(t+1)) = n + 1))$$

   Here the semantics of used symbols such as '(' and '>' and '$\rightarrow$' (implies) are the traditional ones, while '$string2num$' symbolizes a function translating bitstrings into numbers. It is clear that any abstract hardware model can be fully axiomatized in a similar way.
   b) **Reward axioms** defining the computational costs of any hardware instruction, and physical costs of output actions (e.g., control signals

[2] Turing reformulated Gödel's unprovability results in terms of Turing machines (TMs) [56] which subsequently became the most widely used abstract model of computation. It is well-known that there are *universal* TMs that in a certain sense can emulate any other TM or any other known computer. Gödel's integer-based formal language can be used to describe any universal TM, and vice versa.

$y(t)$ encoded in $s(t)$). Related axioms assign values to certain input events (encoded in variable x, a substring of s) representing reward or punishment (e.g., when a Gödel machine-controlled robot bumps into an obstacle). Additional axioms define the total value of the Gödel machine's life as a scalar-valued function of all rewards (e.g., their sum) and costs experienced between cycles 1 and T, etc. For example, assume that $s_{17:18}$ can be changed only through external inputs; the following example axiom says that the total reward increases by 3 whenever such an input equals '11' (unexplained symbols carry the obvious meaning):

$$(\forall t_1 \forall t_2 : [(t_1 < t_2) \wedge (t_1 \geq 1) \wedge (t_2 \leq T) \wedge (s_{17:18}(t_2) = \text{'11'})]$$

$$\rightarrow [R(t_1, t_2) = R(t_1, t_2 - 1) + 3]),$$

where $R(t_1, t_2)$ is interpreted as the cumulative reward between times t_1 and t_2. It is clear that any formal scheme for producing rewards can be fully axiomatized in a similar way.

c) **Environment axioms** restricting the way the environment will produce new inputs (encoded within certain substrings of s) in reaction to sequences of outputs y encoded in s. For example, it may be known in advance that the environment is sampled from an unknown probability distribution that is computable, given the previous history [52, 53, 16], or at least limit-computable [39, 40]. Or, more restrictively, the environment may be some unknown but deterministic computer program [58, 37] sampled from the Speed Prior [41] which assigns low probability to environments that are hard to compute by any method. Or the interface to the environment is Markovian [33], that is, the current input always uniquely identifies the environmental state—a lot of work has been done on this special case [31, 2, 55]. Even more restrictively, the environment may evolve in completely predictable fashion known in advance. All such prior assumptions are perfectly formalizable in an appropriate $\mathcal{A}$ (otherwise we could not write scientific papers about them).

d) **Uncertainty axioms; string manipulation axioms:** Standard axioms for arithmetics and calculus and probability theory [20] and statistics and string manipulation that (in conjunction with the environment axioms) allow for constructing proofs concerning (possibly uncertain) properties of future values of $s(t)$ as well as bounds on expected remaining lifetime / costs / rewards, given some time τ and certain hypothetical values for components of $s(\tau)$ etc. An example theorem saying something about expected properties of future inputs x might look like this:

$$(\forall t_1 \forall \mu \in M : [(1 \leq t_1) \wedge (t_1 + 15597 < T) \wedge (s_{5:9}(t_1) = \text{'01011'})$$

$$\wedge(x_{40:44}(t_1) = \text{`00000'})] \rightarrow (\exists t : [(t_1 < t < t_1 + 15597)$$
$$\wedge(P_\mu(x_{17:22}(t) = \text{`011011'} \mid s(t_1)) > \frac{998}{1000})])),$$

where $P_\mu(. \mid .)$ represents a conditional probability with respect to an axiomatized prior distribution μ from a set of distributions M described by the environment axioms (Item 1c).
Given a particular formalizable hardware (Item 1a) and formalizable assumptions about the possibly probabilistic environment (Item 1c), obviously one can fully axiomatize everything that is needed for proof-based reasoning.

e) **Initial state axioms:** Information about how to reconstruct the initial state $s(1)$ or parts thereof, such that the proof searcher can build proofs including axioms of the type

$$(s_{\mathbf{m:n}}(1) = \mathbf{z}),\ e.g.:\ (s_{7:9}(1) = \text{`010'}).$$

Here and in the remainder of the paper we use bold font in formulas to indicate syntactic place holders (such as **m,n,z**) for symbol strings representing variables (such as *m,n,z*) whose semantics are explained in the text (in the present context z is the bitstring $s_{m:n}(1)$).
Note that it is *no fundamental problem* to fully encode both the hardware description *and* the initial hardware-describing p within p itself. To see this, observe that some software may include a program that can print the software.

f) **Utility axioms** describing the overall goal in the form of utility function u; e.g., equation (1) in Section 2.1.

2. **apply-rule(k, m, n)** takes as arguments the index k (if it exists) of an inference rule such as *modus ponens* (stored in a list of possible inference rules encoded within $p(1)$) and the indices m, n of two previously proven theorems (numbered in order of their creation) in the current *proof*. If applicable, the corresponding inference rule is applied to the addressed theorems and the resulting theorem appended to *proof*. Otherwise the currently tested proof technique is interrupted. This ensures that *proof* is never fed with invalid proofs.
3. **delete-theorem(m)** deletes the m-th theorem in the currently stored *proof*, thus freeing storage such that proof-storing parts of s can be reused and the maximal proof size is not necessarily limited by storage constraints. Theorems deleted from *proof*, however, cannot be addressed any more by *apply-rule* to produce further prolongations of *proof*.
4. **set-switchprog(m,n)** replaces *switchprog* by $s^p_{m:n}$, provided that $s^p_{m:n}$ is indeed a non-empty substring of s^p, the storage writable by proof techniques.
5. **state2theorem(m, n)** takes two integer arguments m, n and tries to transform the current contents of $s_{m:n}$ into a theorem of the form

$$(s_{\mathbf{m:n}}(\mathbf{t_1}) = \mathbf{z}),\ e.g.:\ (s_{6:9}(7775555) = \text{'1001'}),$$

where t_1 represents a time measured (by checking *time*) shortly after *state2theorem* was invoked, and z the bistring $s_{m:n}(t_1)$ (recall the special case $t_1 = 1$ of Item 1e). So we accept the time-labeled current observable contents of any part of s as a theorem that does not have to be proven in an alternative way from, say, the initial state $s(1)$, because the computation so far has already demonstrated that the theorem is true. Thus we may exploit information conveyed by environmental inputs, and the fact that sometimes (but not always) the fastest way to determine the output of a program is to run it.

This non-traditional online interface between syntax and semantics requires special care though. We must avoid inconsistent results through parts of s that change while being read. For example, the present value of a quickly changing instruction pointer *IP* (continually updated by the hardware) may be essentially unreadable in the sense that the execution of the reading subroutine itself will already modify *IP* many times. For convenience, the (typically limited) hardware could be set up such that it stores the contents of fast hardware variables every c cycles in a reserved part of s, such that an appropriate variant of *state2theorem()* could at least translate certain recent values of fast variables into theorems. This, however, will not abolish *all* problems associated with self-observations. For example, the $s_{m:n}$ to be read might also contain the reading procedure's own, temporary, constantly changing string pointer variables, etc.[3] To address such problems on computers with limited memory, *state2theorem* first uses some fixed protocol to check whether the current $s_{m:n}$ is readable at all or whether it might change if it were read by the remaining code of *state2theorem*. If so, or if m, n, are not in the proper range, then the instruction has no further effect. Otherwise it appends an *observed* theorem of the form $(s_{\mathbf{m:n}}(\mathbf{t_1}) = \mathbf{z})$ to *proof*. For example, if the current time is 7770000, then the invocation of *state2theorem(6,9)* might return the theorem $(s_{6:9}(7775555) = \text{'1001'})$, where $7775555 - 7770000 = 5555$ reflects the time needed by *state2theorem* to perform the initial check and to read leading bits off the continually increasing *time* (reading *time* also

[3] We see that certain parts of the current s may not be directly observable without changing the observable itself. Sometimes, however, axioms and previous observations will allow the Gödel machine to *deduce* time-dependent storage contents that are not directly observable. For instance, by analyzing the code being executed through instruction pointer *IP* in the example above, the value of *IP* at certain times may be predictable (or postdictable, after the fact). The values of other variables at given times, however, may not be deducible at all. Such limits of self-observability are reminiscent of Heisenberg's celebrated uncertainty principle [12], which states that certain physical measurements are necessarily imprecise, since the measuring process affects the measured quantity.

costs time) such that it can be sure that 7775555 is a recent proper time label following the start of *state2theorem*.

6. **check()** verifies whether the goal of the proof search has been reached. First it tests whether the last theorem (if any) in *proof* has the form of a **target theorem**. A target theorem states that given the *current* axiomatized utility function u (Item 1f), the utility of a switch from p to the current *switchprog* would be higher than the utility of continuing the execution of p (which would keep searching for alternative *switchprog*s). Target theorems are symbol strings (encoded in s) of the (decoded) form

$$(u[s(\mathbf{t_1}) \oplus (switchbit(\mathbf{t_1}) = \text{`1'}), Env(\mathbf{t_1})] > u[s(\mathbf{t_1}) \oplus (switchbit(\mathbf{t_1}) = \text{`0'}), Env(\mathbf{t_1})]), \quad (2)$$

where the variable t_1 (represented by syntactic place holder $\mathbf{t_1}$) stands for a time step, while all other symbols belong to the alphabet of the theorem-proving calculus, that is, the set of possible target theorems is parameterized only by t_1. Here the calculus should permit the notation $s(\mathbf{t_1}) \oplus (switchbit(\mathbf{t_1}) = \text{`}b\text{'})$ as a shortcut for the state obtained when we replace $switchbit(t_1)$, the true value of the variable bit *switchbit* (encoded in s) at time t_1, by $b \in \{0, 1\}$. This will facilitate the formulation of theorems that compare values conditioned on various alternative hypothetical properties of $s(t_1)$. (Note that $s(t_1)$ may be only partially known by the current proof technique even in environments where $s(t_1)$ and $switchbit(t_1)$ are fully predetermined for all valid t_1.)
The purpose of introducing t_1 is to deal with hardware-specific temporal delays that may be involved in checking and switching—it may take a significant amount of time to match abstract symbol strings found during proof search to the Gödel machine's real current state. If a target theorem has been found, *check()* uses a simple prewired subroutine to check whether there is enough time left to set variable *switchbit* (originally 0) to 1 before the continually increasing *time* will equal t_1. If this subroutine returns a negative result, *check()* exits. Otherwise it sets *switchbit* := 1 (there is no other way of changing *switchbit*). Then it repeatedly tests *time* until $time > t_1$, to make sure the condition of formula (6) was fulfilled at t_1. Then it transfers control to *switchprog* (there is no other way of calling *switchprog*). The *switchprog* may subsequently rewrite all parts of s, excluding hardware-reserved parts such as *time* and x, but including p.

Note that a proof technique does not necessarily have to compute the true expected utilities of switching and not switching—it just needs to determine which is higher. For example, it may be easy to prove that speeding up a subroutine of the proof searcher by a factor of 2 will certainly be worth the negligible (compared to lifetime T) time needed to execute the subroutine-changing algorithm, no matter the precise utility of the switch.

The axiomatic system $\mathcal{A}$ is a defining parameter of a given Gödel machine. Clearly, $\mathcal{A}$ must be strong enough to permit proofs of target theorems. In particular, the theory of uncertainty axioms (Item 1d) must be sufficiently rich. This is no fundamental problem: We simply insert all traditional axioms of probability theory [20].

4 Global Optimality Theorem

Intuitively, at any given time p should execute some self-modification algorithm only if it is the 'best' of all possible self-modifications, given the utility function, which typically depends on available resources, such as storage size and remaining lifetime. At first glance, however, target theorem (6) seems to implicitly talk about just one single modification algorithm, namely, *switchprog*(t_1) as set by the systematic proof searcher at time t_1. Isn't this type of local search greedy? Couldn't it lead to a local optimum instead of a global one? No, it cannot, according to the global optimality theorem:

Theorem 1 (Globally Optimal Self-Changes, given u and $\mathcal{A}$ encoded in p). *Given any formalizable utility function u (Item 1f), and assuming consistency of the underlying formal system $\mathcal{A}$, any self-change of p obtained through execution of some program* switchprog *identified through the proof of a target theorem (6) is globally optimal in the following sense: the utility of starting the execution of the present* switchprog *is higher than the utility of waiting for the proof searcher to produce an alternative* switchprog *later.*

Proof. Target theorem (6) implicitly talks about all the other *switchprog*s that the proof searcher could produce in the future. To see this, consider the two alternatives of the binary decision: (1) either execute the current *switchprog* (set *switchbit* = 1), or (2) keep searching for *proof*s and *switchprog*s (set *switchbit* = 0) until the systematic searcher comes up with an even better *switchprog*. Obviously the second alternative concerns all (possibly infinitely many) potential *switchprog*s to be considered later. That is, if the current *switchprog* were not the 'best', then the proof searcher would not be able to prove that setting *switchbit* and executing *switchprog* will cause higher expected reward than discarding *switchprog*, assuming consistency of $\mathcal{A}$. *Q.E.D.*

4.1 Alternative Relaxed Target Theorem

We may replace the target theorem (6) (Item 6) by the following alternative target theorem:

$$(u[s(\mathbf{t_1}) \oplus (switchbit(\mathbf{t_1}) = \text{'1'}), Env(\mathbf{t_1})] \geq$$
$$u[s(\mathbf{t_1}) \oplus (switchbit(\mathbf{t_1}) = \text{'0'}), Env(\mathbf{t_1})]). \quad (3)$$

The only difference to the original target theorem (6) is that the ">" sign became a "≥" sign. That is, the Gödel machine will change itself as soon as it found a proof that the change will not make things worse. A Global Optimality Theorem similar to Theorem 1 holds.

5 Bias-Optimal Proof Search (BIOPS)

Here we construct a $p(1)$ that is $O()$-optimal in a certain limited sense to be described below, but still might be improved as it is not necessarily optimal in the sense of the given u (for example, the u of equation (1) neither mentions nor cares for $O()$-optimality). Our Bias-Optimal Proof Search (BIOPS) is essentially an application of Universal Search [23, 25] to proof search. Previous practical variants and extensions of universal search have been applied [36, 38, 50, 45] to *offline* program search tasks where the program inputs are fixed such that the same program always produces the same results. In our *online* setting, however, BIOPS has to take into account that the same proof technique started at different times may yield different proofs, as it may read parts of s (e.g., inputs) that change as the machine's life proceeds.

BIOPS starts with a probability distribution P (the initial bias) on the proof techniques w that one can write in $\mathcal{L}$, e.g., $P(w) = K^{-l(w)}$ for programs composed from K possible instructions [25]. BIOPS is *near-bias-optimal* [45] in the sense that it will not spend much more time on any proof technique than it deserves, according to its probabilistic bias, namely, not much more than its probability times the total search time:

Definition 1 (Bias-Optimal Searchers [45]). Let $\mathcal{R}$ be a problem class, $\mathcal{C}$ be a search space of solution candidates (where any problem $r \in \mathcal{R}$ should have a solution in $\mathcal{C}$), $P(q \mid r)$ be a task-dependent bias in the form of conditional probability distributions on the candidates $q \in \mathcal{C}$. Suppose that we also have a predefined procedure that creates and tests any given q on any $r \in \mathcal{R}$ within time $t(q, r)$ (typically unknown in advance). Then *a searcher is n-bias-optimal ($n \geq 1$) if for any maximal total search time $T_{total} > 0$ it is guaranteed to solve any problem $r \in \mathcal{R}$ if it has a solution $p \in \mathcal{C}$ satisfying $t(p, r) \leq P(p \mid r)\ T_{total}/n$. It is bias-optimal if $n = 1$.*

Method 5.1 (BIOPS) In phase $(i = 1, 2, 3, \ldots)$ Do: For all self-delimiting [25] proof techniques $w \in \mathcal{L}$ satisfying $P(w) \geq 2^{-i}$ Do:

1. Run w until halt or error (such as division by zero) or $2^i P(w)$ steps consumed.
2. Undo effects of w on s^p (does not cost significantly more time than executing w).

A proof technique w can interrupt Method 5.1 only by invoking instruction *check()* (Item 6), which may transfer control to *switchprog* (which possibly

even will delete or rewrite Method 5.1). Since the initial p runs on the formalized hardware, and since proof techniques tested by p can read p and other parts of s, they can produce proofs concerning the (expected) performance of p and BIOPS itself. Method 5.1 at least has the optimal *order* of computational complexity in the following sense.

Theorem 2. *If independently of variable* time(s) *some unknown fast proof technique w would require at most $f(k)$ steps to produce a proof of difficulty measure k (an integer depending on the nature of the task to be solved), then Method 5.1 will need at most $O(f(k))$ steps.*

Proof. It is easy to see that Method 5.1 will need at most $O(f(k)/P(w)) = O(f(k))$ steps—the constant factor $1/P(w)$ does not depend on k. *Q.E.D.*

Note again, however, that the proofs themselves may concern quite different, arbitrary formalizable notions of optimality (stronger than those expressible in the $O()$-notation) embodied by the given, problem-specific, formalized utility function u. This may provoke useful, constant-affecting rewrites of the initial proof searcher despite its limited (yet popular and widely used) notion of $O()$-optimality.

5.1 How a Surviving Proof Searcher May Use BIOPS to Solve Remaining Proof Search Tasks

The following is not essential for this chapter. Let us assume that the execution of the *switchprog* corresponding to the first found target theorem has not rewritten the code of p itself—the current p is still equal to $p(1)$—and has reset *switchbit* and returned control to p such that it can continue where it was interrupted. In that case the BIOPS subroutine of $p(1)$ can use the Optimal Ordered Problem Solver OOPS [45] to accelerate the search for the n-th target theorem ($n > 1$) by reusing proof techniques for earlier found target theorems where possible. The basic ideas are as follows (details: [45]).

Whenever a target theorem has been proven, $p(1)$ *freezes* the corresponding proof technique: its code becomes non-writable by proof techniques to be tested in later proof search tasks. But it remains readable, such that it can be copy-edited and/or invoked as a subprogram by future proof techniques. We also allow prefixes of proof techniques to temporarily rewrite the probability distribution on their suffixes [45], thus essentially rewriting the probability-based search procedure (an incremental extension of Method 5.1) based on previous experience. As a side-effect we metasearch for faster search procedures, which can greatly accelerate the learning of new tasks [45].

Given a new proof search task, BIOPS performs OOPS by spending half the total search time on a variant of Method 5.1 that searches only among self-delimiting [24, 6] proof techniques starting with the most recently frozen proof technique. The rest of the time is spent on fresh proof techniques with arbitrary prefixes (which may reuse previously frozen proof techniques though) [45]. (We could also search for a *generalizing* proof technique solving all proof

search tasks so far. In the first half of the search we would not have to test proof techniques on tasks other than the most recent one, since we already know that their prefixes solve the previous tasks [45].)

It can be shown that OOPS is essentially *8-bias-optimal* (see Def. 1), given either the initial bias or intermediate biases due to frozen solutions to previous tasks [45]. This result immediately carries over to BIOPS. To summarize, BIOPS essentially allocates part of the total search time for a new task to proof techniques that exploit previous successful proof techniques in computable ways. If the new task can be solved faster by copy-editing / invoking previously frozen proof techniques than by solving the new proof search task from scratch, then BIOPS will discover this and profit thereof. If not, then at least it will not be significantly slowed down by the previous solutions—BIOPS will remain 8-bias-optimal.

Recall, however, that BIOPS is not the only possible way of initializing the Gödel machine's proof searcher.

6 Discussion & Additional Relations to Previous Work

Here we list a few examples of possible types of self-improvements (Sect. 6.1), Gödel machine applicability to various tasks defined by various utility functions and environments (Sect. 6.2), probabilistic hardware (Sect. 6.3), and additional relations to previous work (Sect. 6.4). We also briefly discuss self-reference and consciousness (Sect. 6.6), and provide a list of answers to frequently asked questions (Sect. 6.7).

6.1 Possible Types of Gödel Machine Self-improvements

Which provably useful self-modifications are possible? There are few limits to what a Gödel machine might do:

1. In one of the simplest cases it might leave its basic proof searcher intact and just change the ratio of time-sharing between the proof searching subroutine and the subpolicy e—those parts of p responsible for interaction with the environment.
2. Or the Gödel machine might modify e only. For example, the initial e may regularly store limited memories of past events somewhere in s; this might allow p to derive that it would be useful to modify e such that e will conduct certain experiments to increase the knowledge about the environment, and use the resulting information to increase reward intake. In this sense the Gödel machine embodies a principled way of dealing with the exploration versus exploitation problem [19]. Note that the *expected* utility of conducting some experiment may exceed the one of not conducting it, even when the experimental outcome later suggests to keep acting in line with the previous e.

3. The Gödel machine might also modify its very axioms to speed things up. For example, it might find a proof that the original axioms should be replaced or augmented by theorems derivable from the original axioms.
4. The Gödel machine might even change its own utility function and target theorem, but can do so only if their *new* values are provably better according to the *old* ones.
5. In many cases we do not expect the Gödel machine to replace its proof searcher by code that completely abandons the search for proofs. Instead, we expect that only certain subroutines of the proof searcher will be sped up—compare the example at the end of Item 6 in Section 3.1—or that perhaps just the order of generated proofs will be modified in problem-specific fashion. This could be done by modifying the probability distribution on the proof techniques of the initial bias-optimal proof searcher from Section 5.
6. Generally speaking, the utility of limited rewrites may often be easier to prove than the one of total rewrites. For example, suppose it is 8:00 PM and our Gödel machine-controlled agent's permanent goal is to maximize future expected reward, using the (alternative) target theorem (4.1). Part thereof is to avoid hunger. There is nothing in its fridge, and shops close down at 8:30 PM. It does not have time to optimize its way to the supermarket in every little detail, but if it does not get going right now it will stay hungry tonight (in principle such near-future consequences of actions should be easily provable, possibly even in a way related to how humans prove advantages of potential actions to themselves). That is, if the agent's previous policy did not already include, say, an automatic daily evening trip to the supermarket, the policy provably should be rewritten at least limitedly and simply right now, while there is still time, such that the agent will surely get some food tonight, without affecting less urgent future behavior that can be optimized/decided later, such as details of the route to the food, or of tomorrow's actions.
7. In certain uninteresting environments reward is maximized by becoming dumb. For example, a given task may require to repeatedly and forever execute the same pleasure center-activating action, as quickly as possible. In such cases the Gödel machine may delete most of its more time-consuming initial software including the proof searcher.
8. Note that there is no reason why a Gödel machine should not augment its own hardware. Suppose its lifetime is known to be 100 years. Given a hard problem and axioms restricting the possible behaviors of the environment, the Gödel machine might find a proof that its expected cumulative reward will increase if it invests 10 years into building faster computational hardware, by exploiting the physical resources of its environment.

6.2 Example Applications

Example 1 (Maximizing expected reward with bounded resources). A robot that needs at least 1 liter of gasoline per hour interacts with a partially unknown environment, trying to find hidden, limited gasoline depots to occasionally refuel its tank. It is rewarded in proportion to its lifetime, and dies after at most 100 years or as soon as its tank is empty or it falls off a cliff, etc. The probabilistic environmental reactions are initially unknown but assumed to be sampled from the axiomatized Speed Prior [41], according to which hard-to-compute environmental reactions are unlikely. This permits a computable strategy for making near-optimal predictions [41]. One by-product of maximizing expected reward is to maximize expected lifetime.

Less general, more traditional examples that do not involve significant interaction with a probabilistic environment are also easily dealt with in the reward-based framework:

Example 2 (Time-limited NP-hard optimization). The initial input to the Gödel machine is the representation of a connected graph with a large number of nodes linked by edges of various lengths. Within given time T it should find a cyclic path connecting all nodes. The only real-valued reward will occur at time T. It equals 1 divided by the length of the best path found so far (0 if none was found). There are no other inputs. The by-product of maximizing expected reward is to find the shortest path findable within the limited time, given the initial bias.

Example 3 (Fast theorem proving). Prove or disprove as quickly as possible that all even integers > 2 are the sum of two primes (Goldbach's conjecture). The reward is $1/t$, where t is the time required to produce and verify the first such proof.

Example 4 (Optimize any suboptimal problem solver). Given any formalizable problem, implement a suboptimal but known problem solver as software on the Gödel machine hardware, and let the proof searcher of Section 5 run in parallel.

6.3 Probabilistic Gödel Machine Hardware

Above we have focused on an example deterministic machine. It is straightforward to extend this to computers whose actions are computed in probabilistic fashion, given the current state. Then the expectation calculus used for probabilistic aspects of the environment simply has to be extended to the hardware itself, and the mechanism for verifying proofs has to take into account that there is no such thing as a certain theorem—at best there are formal statements which are true with such and such probability. In fact, this may be the most realistic approach as any physical hardware is error-prone, which should be taken into account by realistic probabilistic Gödel machines.

Probabilistic settings also automatically avoid certain issues of axiomatic consistency. For example, predictions proven to come true with probability less than 1.0 do not necessarily cause contradictions even when they do not match the observations.

6.4 More Relations to Previous Work on Less General Self-improving Machines

Despite (or maybe because of) the ambitiousness and potential power of self-improving machines, there has been little work in this vein outside our own labs at IDSIA and TU Munich. Here we will list essential differences between the Gödel machine and our previous approaches to 'learning to learn,' 'metalearning,' self-improvement, self-optimization, etc.

1. **Gödel Machine versus Success-Story Algorithm and Other Metalearners**
   A learner's modifiable components are called its policy. An algorithm that modifies the policy is a learning algorithm. If the learning algorithm has modifiable components represented as part of the policy, then we speak of a self-modifying policy (SMP) [48]. SMPs can modify the way they modify themselves etc. The Gödel machine has an SMP.
   In previous work we used the *success-story algorithm* (SSA) to force some (stochastic) SMPs to trigger better and better self-modifications [35, 49, 48, 50]. During the learner's life-time, SSA is occasionally called at times computed according to SMP itself. SSA uses backtracking to undo those SMP-generated SMP-modifications that have not been empirically observed to trigger lifelong reward accelerations (measured up until the current SSA call—this evaluates the long-term effects of SMP-modifications setting the stage for later SMP-modifications). SMP-modifications that survive SSA represent a lifelong success history. Until the next SSA call, they build the basis for additional SMP-modifications. Solely by self-modifications our SMP/SSA-based learners solved a complex task in a partially observable environment whose state space is far bigger than most found in the literature [48].
   The Gödel machine's training algorithm is theoretically more powerful than SSA though. SSA empirically measures the usefulness of previous self-modifications, and does not necessarily encourage provably optimal ones. Similar drawbacks hold for Lenat's human-assisted, non-autonomous, self-modifying learner [22], our Meta-Genetic Programming [32] extending Cramer's Genetic Programming [8, 1], our metalearning economies [32] extending Holland's machine learning economies [15], and gradient-based metalearners for continuous program spaces of differentiable recurrent neural networks [34, 13]. All these methods, however, could be used to seed $p(1)$ with an initial policy.

2. **Gödel Machine versus** OOPS **and** OOPS-RL
The Optimal Ordered Problem Solver OOPS [45, 42] (used by BIOPS in Sect. 5.1) is a bias-optimal (see Def. 1) way of searching for a program that solves each problem in an ordered sequence of problems of a reasonably general type, continually organizing and managing and reusing earlier acquired knowledge. Solomonoff recently also proposed related ideas for a *scientist's assistant* [54] that modifies the probability distribution of universal search [23] based on experience.
As pointed out earlier [45] (section on OOPS limitations), however, OOPS-like methods are not directly applicable to general lifelong reinforcement learning (RL) tasks [19] such as those for which AIXI [16] was designed. The simple and natural but limited optimality notion of OOPS is *bias-optimality* (Def. 1): OOPS is a near-bias-optimal searcher for programs which compute solutions that one can quickly verify (costs of verification are taken into account). For example, one can quickly test whether some currently tested program has computed a solution to the *towers of Hanoi* problem used in the earlier paper [45]: one just has to check whether the third peg is full of disks.
But general RL tasks are harder. Here, in principle, the evaluation of the value of some behavior consumes the learner's entire life! That is, the naive test of whether a program is good or not would consume the entire life. That is, we could test only one program; afterwards life would be over.
So general RL machines need a more general notion of optimality, and must do things that plain OOPS does not do, such as predicting *future* tasks and rewards. It is possible to use two OOPS -modules as components of a rather general reinforcement learner (OOPS-RL), one module learning a predictive model of the environment, the other one using this *world model* to search for an action sequence maximizing expected reward [45, 44]. Despite the bias-optimality properties of OOPS for certain ordered task sequences, however, OOPS-RL is not necessarily the best way of spending limited computation time in general RL situations.
A provably optimal RL machine must somehow *prove* properties of otherwise un-testable behaviors (such as: what is the expected reward of this behavior which one cannot naively test as there is not enough time). That is part of what the Gödel machine does: It tries to greatly cut testing time, replacing naive time-consuming tests by much faster proofs of predictable test outcomes whenever this is possible.
Proof verification itself can be performed very quickly. In particular, verifying the correctness of a found proof typically does not consume the remaining life. Hence the Gödel machine may use OOPS as a bias-optimal proof-searching submodule. Since the proofs themselves may concern quite different, *arbitrary* notions of optimality (not just bias-optimality), the Gödel machine is more general than plain OOPS. But it is not just an extension of OOPS. Instead of OOPS it may as well use non-bias-optimal alternative methods to initialize its proof searcher. On the other hand, OOPS

is not just a precursor of the Gödel machine. It is a stand-alone, incremental, bias-optimal way of allocating runtime to programs that reuse previously successful programs, and is applicable to many traditional problems, including but not limited to proof search.

3. **Gödel Machine versus** AIXI **etc.**
Unlike Gödel machines, Hutter's recent AIXI *model* [16] generally needs *unlimited* computational resources per input update. It combines Solomonoff's universal prediction scheme [52, 53] with an *expectimax* computation. In discrete cycle $k = 1, 2, 3, \ldots$, action $y(k)$ results in perception $x(k)$ and reward $r(k)$, both sampled from the unknown (reactive) environmental probability distribution μ. AIXI defines a mixture distribution ξ as a weighted sum of distributions $\nu \in \mathcal{M}$, where $\mathcal{M}$ is any class of distributions that includes the true environment μ. For example, $\mathcal{M}$ may be a sum of all computable distributions [52, 53], where the sum of the weights does not exceed 1. In cycle $k + 1$, AIXI selects as next action the first in an action sequence maximizing ξ-predicted reward up to some given horizon. Recent work [18] demonstrated AIXI 's optimal use of observations as follows. The Bayes-optimal policy p^ξ based on the mixture ξ is self-optimizing in the sense that its average utility value converges asymptotically for all $\mu \in \mathcal{M}$ to the optimal value achieved by the (infeasible) Bayes-optimal policy p^μ which knows μ in advance. The necessary condition that $\mathcal{M}$ admits self-optimizing policies is also sufficient. Furthermore, p^ξ is Pareto-optimal in the sense that there is no other policy yielding higher or equal value in *all* environments $\nu \in \mathcal{M}$ and a strictly higher value in at least one [18].
While AIXI clarifies certain theoretical limits of machine learning, it is computationally intractable, especially when $\mathcal{M}$ includes all computable distributions. This drawback motivated work on the time-bounded, asymptotically optimal AIXI*(t,l)* system [16] and the related HSEARCH [17], both already discussed in Section 2.4, which also lists the advantages of the Gödel machine. Both methods, however, could be used to seed the Gödel machine with an *initial* policy.
It is the *self-referential* aspects of the Gödel machine that relieve us of much of the burden of careful algorithm design required for AIXI*(t,l)* and HSEARCH. They make the Gödel machine both conceptually simpler *and* more general than AIXI*(t,l)* and HSEARCH.

6.5 Are Humans Probabilistic Gödel Machines?

We do not know. We think they better be. Their initial underlying formal system for dealing with uncertainty seems to differ substantially from those of traditional expectation calculus and logic though—compare Items 1c and 1d in Sect. 3.1 as well as the supermarket example in Sect. 6.1.

6.6 Gödel Machines and Consciousness

In recent years the topic of consciousness has gained some credibility as a serious research issue, at least in philosophy and neuroscience, e.g., [9]. However, there is a lack of *technical* justifications of consciousness: so far nobody has shown that consciousness is really useful for solving problems, although problem solving is considered of central importance in philosophy [30].

The fully self-referential Gödel machine may be viewed as providing just such a technical justification. It is "conscious" or "self-aware" in the sense that its entire behavior is open to self-introspection, and modifiable. It may "step outside of itself" [14] by executing self-changes that are provably good, where the proof searcher itself is subject to analysis and change through the proof techniques it tests. And this type of total self-reference is precisely the reason for its optimality as a problem solver in the sense of Theorem 1.

6.7 Frequently Asked Questions

In the past half year the author frequently fielded questions about the Gödel machine. Here a list of answers to typical questions.

1. **Q:** *Does the exact business of formal proof search really make sense in the uncertain real world?*
   A: Yes, it does. We just need to insert into $p(1)$ the standard axioms for representing uncertainty and for dealing with probabilistic settings and expected rewards etc. Compare items 1d and 1c in Section 3.1, and the definition of utility as an *expected* value in equation (1).
2. **Q:** *The target theorem (6) seems to refer only to the* very first *self-change, which may completely rewrite the proof-search subroutine—doesn't this make the proof of Theorem 1 invalid? What prevents later self-changes from being destructive?*
   A: This is fully taken care of. Please look once more at the proof of Theorem 1, and note that the first self-change will be executed only if it is provably useful (in the sense of the present untility function u) for all future self-changes (for which the present self-change is setting the stage). This is actually the main point of the whole Gödel machine set-up.
3. **Q** (related to the previous item): *The Gödel machine implements a meta-learning behavior: what about a meta-meta, and a meta-meta-meta level?*
   A: The beautiful thing is that all meta-levels are automatically collapsed into one: any proof of a target theorem automatically proves that the corresponding self-modification is good for all further self-modifications affected by the present one, in recursive fashion.
4. **Q:** *The Gödel machine software can produce only computable mappings from input sequences to output sequences. What if the environment is non-computable?*
   A: Many physicists and other scientists (exceptions: [58, 37]) actually do assume the real world makes use of all the real numbers, most of which

are incomputable. Nevertheless, theorems and proofs are just finite symbol strings, and all treatises of physics contain only computable axioms and theorems, even when some of the theorems can be interpreted as making statements about uncountably many objects, such as all the real numbers. (Note though that the Löwenheim-Skolem Theorem [27, 51] implies that any first order theory with an uncountable model such as the real numbers also has a countable model.) Generally speaking, formal descriptions of non-computable objects do *not at all* present a fundamental problem—they may still allow for finding a strategy that provably maximizes utility. If so, a Gödel machine can exploit this. If not, then humans will not have a fundamental advantage over Gödel machines.

5. **Q:** *Isn't automated theorem-proving very hard? Current AI systems cannot prove nontrivial theorems without human intervention at crucial decision points.*
   A: More and more important mathematical proofs (four color theorem, etc.) heavily depend on automated proof search. And traditional theorem provers do not even make use of our novel notions of proof techniques and $O()$-optimal proof search. Of course, some proofs are indeed hard to find, but here humans and Gödel machines face the same fundamental limitations.
6. **Q:** *Don't the "no free lunch theorems" [57] say that it is impossible to construct universal problem solvers?*
   A: No, they do not. They refer to the very special case of problems sampled from *i.i.d.* uniform distributions on *finite* problem spaces. See the discussion of no free lunch theorems in an earlier paper [45].
7. **Q:** *Can't the Gödel machine switch to a program* switchprog *that rewrites the utility function to a "bogus" utility function that makes unfounded promises of big rewards in the near future?*
   A: No, it cannot. It should be obvious that rewrites of the utility function can happen only if the Gödel machine first can prove that the rewrite is useful according to the *present* utility function.

7 Conclusion

The initial software $p(1)$ of our machine runs an initial problem solver, e.g., one of Hutter's approaches [17, 16] which have at least an optimal *order* of complexity. Simultaneously, it runs an $O()$-optimal initial proof searcher using an online variant of Universal Search to test *proof techniques*, which are programs able to compute proofs concerning the system's own future performance, based on an axiomatic system $\mathcal{A}$ encoded in $p(1)$, describing a formal *utility* function u, the hardware and $p(1)$ itself. If there is no provably good, globally optimal way of rewriting $p(1)$ at all, then humans will not find one either. But if there is one, then $p(1)$ itself can find and exploit it. This approach

yields the first class of theoretically sound, fully self-referential, optimally efficient, general problem solvers.

After the theoretical discussion in Sects. 1 through 5, one practical question remains: to build a particular, especially practical Gödel machine with small initial constant overhead, which generally useful theorems should one add as axioms to $\mathcal{A}$ (as initial bias) such that the initial searcher does not have to prove them from scratch?

8 Acknowledgments

Thanks to Alexey Chernov, Marcus Hutter, Jan Poland, Sepp Hochreiter, Ray Solomonoff, Leonid Levin, Shane Legg, Alex Graves, Matteo Gagliolo, Viktor Zhumatiy, Ben Goertzel, Will Pearson, Faustino Gomez, and many others for useful comments on drafts or summaries or earlier versions of this paper.

References

1. Banzhaf W, Nordin P, Keller RE, Francone FD (1998) *Genetic Programming – An Introduction*. Morgan Kaufmann Publishers, San Francisco, CA.
2. Bellman R (1961) *Adaptive Control Processes*. Princeton University Press, Princeton, NJ.
3. Blum M (1967) A machine-independent theory of the complexity of recursive functions. *Journal of the ACM*, 14(2):322–336.
4. Blum M On effective procedures for speeding up algorithms. *Journal of the ACM*, 18(2):290–305.
5. Cantor G Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen. *Crelle's Journal für Mathematik*, 77:258–263.
6. Chaitin GJ (1975) A theory of program size formally identical to information theory. *Journal of the ACM*, 22:329–340.
7. Clocksin WF,Mellish CS (1987) *Programming in Prolog*. Springer, Berlin, 3rd edition.
8. Cramer NL (1985) A representation for the adaptive generation of simple sequential programs. In Grefenstette JJ (ed) *Proceedings of an International Conference on Genetic Algorithms and Their Applications, Carnegie-Mellon University, July 24-26, 1985*, Lawrence Erlbaum, Hillsdale, NJ.
9. Crick F, Koch C (1998) Consciousness and neuroscience. *Cerebral Cortex*, 8:97–107.
10. Fitting MC (1996) *First-Order Logic and Automated Theorem Proving*. Graduate Texts in Computer Science. Springer, Berlin, 2nd edition.
11. Gödel K (1931) Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38:173–198.
12. Heisenberg W (1925) Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 33:879–893.

13. Hochreiter S, Younger AS, Conwell PR (2001) Learning to learn using gradient descent. In *Proc. Intl. Conf. on Artificial Neural Networks (ICANN-2001)*, volume 2130 of *LLCS* Springer, Berlin, Heidelberg.
14. Hofstadter D (!979) *Gödel, Escher, Bach: an Eternal Golden Braid*. Basic Books, New York.
15. Holland JH (1975) Properties of the bucket brigade. In *Proceedings of an International Conference on Genetic Algorithms*. Lawrence Erlbaum, Hillsdale, NJ.
16. Hutter M (2001) Towards a universal theory of artificial intelligence based on algorithmic probability and sequential decisions. *Proceedings of the 12th European Conference on Machine Learning (ECML-2001)*.
17. Hutter M (2002) The fastest and shortest algorithm for all well-defined problems. *International Journal of Foundations of Computer Science*, 13(3):431–443.
18. Hutter M (2002) Self-optimizing and Pareto-optimal policies in general environments based on Bayes-mixtures. In *Proc. 15th Annual Conf. on Computational Learning Theory (COLT 2002)*, volume 2375 of *LNAI*, Springer, Berlin.
19. Kaelbling LP, Littman ML, Moore AW Reinforcement learning: a survey. *Journal of AI research*, 4:237–285.
20. Kolmogorov AN (1933) *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933.
21. Kolmogorov AN (1965) Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1:1–11.
22. Lenat D (1983) Theory formation by heuristic search. *Machine Learning*, 21.
23. Levin LA (1973) Universal sequential search problems. *Problems of Information Transmission*, 9(3):265–266.
24. Levin LA (1974) Laws of information (nongrowth) and aspects of the foundation of probability theory. *Problems of Information Transmission*, 10(3):206–210.
25. Levin LA (1984) Randomness conservation inequalities: Information and independence in mathematical theories. *Information and Control*, 61:15–37.
26. Li M,Vitányi PMB (!997) *An Introduction to Kolmogorov Complexity and its Applications*. Springer, Berlin, 2nd edition.
27. Löwenheim L (1915) Über Möglichkeiten im Relativkalkül. *Mathematische Annalen*, 76:447–470.
28. Moore CH, Leach GC (1970) FORTH - a language for interactive computing, 1970. `http://www.ultratechnology.com`.
29. Penrose R (1994) *Shadows of the mind*. Oxford University Press, Oxford.
30. Popper KR (1999) *All Life Is Problem Solving*. Routledge, London.
31. Samuel AL (1959) Some studies in machine learning using the game of checkers. *IBM Journal on Research and Development*, 3:210–229.
32. Schmidhuber J (1987) Evolutionary principles in self-referential learning. Diploma thesis, Institut für Informatik, Technische Universität München.
33. Schmidhuber J (1991) Reinforcement learning in Markovian and non-Markovian environments. In Lippman DS, Moody JE, Touretzky DS (eds) *Advances in Neural Information Processing Systems 3*, Morgan Kaufmann, Los Altos, CA.
34. Schmidhuber J A self-referential weight matrix. In *Proceedings of the International Conference on Artificial Neural Networks, Amsterdam*, Springer, Berlin.

35. Schmidhuber J (1994) On learning how to learn learning strategies. Technical Report FKI-198-94, Fakultät für Informatik, Technische Universität München, 1994. See [50, 48].
36. Schmidhuber J (1995) Discovering solutions with low Kolmogorov complexity and high generalization capability. In Prieditis A and Russell S (eds) *Machine Learning: Proceedings of the Twelfth International Conference*. Morgan Kaufmann, San Francisco, CA.
37. Schmidhuber J (1997) A computer scientist's view of life, the universe, and everything. In Freksa C, Jantzen M, Valk R (eds) *Foundations of Computer Science: Potential - Theory - Cognition*, volume 1337 of *LLNCS*, Springer, Berlin.
38. Schmidhuber J (1997) Discovering neural nets with low Kolmogorov complexity and high generalization capability. *Neural Networks*, 10(5):857–873.
39. Schmidhuber J (2000) Algorithmic theories of everything. Technical Report IDSIA-20-00, quant-ph/0011122, IDSIA. Sections 1-5: see [40]; Section 6: see [41].
40. Schmidhuber J (2002) Hierarchies of generalized Kolmogorov complexities and nonenumerable universal measures computable in the limit. *International Journal of Foundations of Computer Science*, 13(4):587–612.
41. Schmidhuber J (2002) The Speed Prior: a new simplicity measure yielding near-optimal computable predictions. In Kivinen J, Sloan RH (eds) *Proceedings of the 15th Annual Conference on Computational Learning Theory (COLT 2002)*, Lecture Notes in Artificial Intelligence, Springer, Berlin.
42. Schmidhuber J (2003) Bias-optimal incremental problem solving. In Becker S, Thrun S, Obermayer K (eds) *Advances in Neural Information Processing Systems 15*, MIT Press, Cambridge, MA.
43. Schmidhuber J (2003) Gödel machines: self-referential universal problem solvers making provably optimal self-improvements. Technical Report IDSIA-19-03, arXiv:cs.LO/0309048 v2, IDSIA.
44. J. Schmidhuber. The new AI: General & sound & relevant for physics. In this volume.
45. Schmidhuber J (2004) Optimal ordered problem solver. *Machine Learning*, 54:211–254.
46. Schmidhuber J (2005) Gödel machines: Towards a Technical Justification of Consciousness. In Kudenko D, Kazakov D, Alonso E (eds) *Adaptive Agents and Multi-Agent Systems III*, LNCS 3394, Springer, Berlin.
47. Schmidhuber J (2005) Completely Self-Referential Optimal Reinforcement Learners. In Duch W et al (eds) *Proc. Intl. Conf. on Artificial Neural Networks ICANN'05*, LNCS 3697, Springer, Berlin, Heidelberg.
48. Schmidhuber J, Zhao J, Schraudolph N (1997) Reinforcement learning with self-modifying policies. In Thrun S, Pratt L (eds) *Learning to learn*, Kluwer, Norwell, MA.
49. Schmidhuber J, Zhao J, Wiering M (1996) Simple principles of metalearning. Technical Report IDSIA-69-96, IDSIA. See [50, 48].
50. Schmidhuber J, Zhao J, Wiering M (1997) Shifting inductive bias with success-story algorithm, adaptive Levin search, and incremental self-improvement. *Machine Learning*, 28:105–130.
51. Skolem T (1919) Logisch-kombinatorische Untersuchungen über Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theorem über dichte Mengen. *Skrifter utgit av Videnskapsselskapet in Kristiania, I, Mat.-Nat. Kl.*, N4:1–36.

52. Solomonoff R (1964) A formal theory of inductive inference. Part I. *Information and Control*, 7:1–22.
53. Solomonoff R (1978) Complexity-based induction systems. *IEEE Transactions on Information Theory*, IT-24(5):422–432.
54. Solomonoff R (2003) Progress in incremental machine learning—Preliminary Report for NIPS 2002 Workshop on Universal Learners and Optimal Search; revised Sept 2003. Technical Report IDSIA-16-03, IDSIA.
55. Sutton R, Barto A (1998) *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA.
56. Turing A (1936) On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, Series 2*, 41:230–267.
57. Wolpert DH, Macready DG (1997) No free lunch theorems for search. *IEEE Transactions on Evolutionary Computation*, 1.
58. Zuse K (1969) *Rechnender Raum*. Friedrich Vieweg & Sohn, Braunschweig. English translation: *Calculating Space,* MIT Technical Translation AZT-70-164-GEMIT, MIT (Proj. MAC), Cambridge, MA.