Artificial Scientists & Artists Based on the Formal Theory of Creativity

Jürgen Schmidhuber
IDSIA, Galleria 2, 6928 Manno-Lugano, Switzerland
University of Lugano & SUPSI, Switzerland

Abstract

I have argued that a simple but general formal theory of creativity based on reward for creating or finding *novel patterns allowing for data compression progress* explains many essential aspects of intelligence including science, art, music, humor. Here I discuss what kind of general bias towards algorithmic regularities we insert into our robots by implementing the principle, why that bias is good, and how the approach greatly generalizes the field of *active learning*. I provide discrete and continuous time formulations for ongoing work on building an Artificial General Intelligence (AGI) based on variants of the artificial creativity framework.

Introduction

Since 1990 I have built artificial agents with an intrinsic desire to create / discover more *novel patterns*, that is, data predictable or compressible in hitherto unknown ways (Sch91b; Sch91a; SHS95; Sch97c; Sch02a; Sch06a; Sch07; Sch09c; Sch09b; Sch09a). The agents embody approximations of a simple, but general, formal theory of creativity explaining essential aspects of human or non-human intelligence including science, art, music, humor (Sch06a; Sch07; Sch09c; Sch09b; Sch09a). Crucial ingredients are: (1) A predictor or compressor of the continually growing history of actions and sensory inputs, reflecting what's currently known about how the world works, (2) A learning algorithm that continually improves the predictor or compressor (detecting novel spatio-temporal patterns that subsequently become known patterns), (3) Intrinsic rewards measuring the predictor's or compressor's improvements due to the learning algorithm, (4) A reward optimizer or reinforcement learner, which translates those rewards into action sequences expected to optimize future reward, thus motivating the agent to create additional novel patterns predictable or compressible in previously unknown ways. We implemented the following variants: (A) Intrinsic reward as measured by improvement in mean squared prediction error (1991) (Sch91a), (B) Intrinsic reward as measured by relative entropies between the agent's priors and posteriors (1995) (SHS95), (C) Learning of probabilistic, hierarchical programs and skills through zero-sum intrinsic reward games of two players, each trying to out-predict or surprise the other, taking into account the computational costs of learning, and learning *when* to learn and *what* to learn (1997-2002) (Sch02a). (A, B, C) also showed experimentally how intrinsic rewards can substantially accelerate goal-directed learning and external reward intake. We also discussed (D) Mathematically optimal, intrinsically motivated systems driven by prediction progress or compression progress (2006-2009) (Sch06a; Sch07; Sch09c; Sch09b).

How does the compression progress drive explain, say, **humor**? Consider the following statement: *Biological organisms are driven by the "Four Big F's": Feeding, Fighting, Fleeing, Sexual Activity.* Some subjective observers who read this for the first time think it is funny. Why? As the eyes are sequentially scanning the text the brain receives a complex visual input stream. The latter is subjectively partially compressible as it relates to the observer's previous knowledge about letters and words. That is, given the reader's current knowledge and current compressor, the raw data can be encoded by fewer bits than required to store random data of the same size. But the punch line after the last comma is unexpected for those who expected another "F". Initially this failed expectation results in sub-optimal data compression—storage of expected events does not cost anything, but deviations from predictions require extra bits to encode them. The compressor, however, does not stay the same forever: within a short time interval its learning algorithm improves its performance on the data seen so far, by discovering the non-random, non-arbitrary and therefore compressible pattern relating the punch line to previous text and previous knowledge about the "Four Big F's." This saves a few bits of storage. The number of saved bits (or a similar measure of learning progress) becomes the observer's intrinsic reward, possibly strong enough to motivate him to read on in search for more reward through additional yet unknown patterns. The recent joke, however, will never be novel or funny again.

How does the theory informally explain the motivation to create or perceive **art and music** (Sch97b; Sch97a; Sch06a; Sch07; Sch09c; Sch09b; Sch09a)? For

example, why are some melodies more interesting or aesthetically rewarding than others? Not the one the listener (composer) just heard (played) twenty times in a row. It became too subjectively predictable in the process. Not the weird one with completely unfamiliar rhythm and tonality. It seems too irregular and contain too much arbitrariness and subjective noise. The observer (creator) of the data is interested in melodies that are unfamiliar enough to contain somewhat unexpected harmonies or beats etc., but familiar enough to allow for quickly recognizing the presence of a new learnable regularity or compressibility in the sound stream: a novel pattern! Sure, it will get boring over time, but not yet. All of this perfectly fits our principle: The current compressor of the observer or data creator tries to compress his history of acoustic and other inputs where possible. The action selector tries to find history-influencing actions such that the continually growing historic data allows for improving the compressor's performance. The interesting or aesthetically rewarding musical and other subsequences are precisely those with previously unknown yet learnable types of regularities, because they lead to compressor improvements. The boring patterns are those that are either already perfectly known or arbitrary or random, or whose structure seems too hard to understand. Similar statements not only hold for other dynamic art including film and dance (take into account the compressibility of action sequences), but also for "static" art such as painting and sculpture, created through action sequences of the artist, and perceived as dynamic spatio-temporal patterns through active attention shifts of the observer. When not occupied with optimizing *external* reward, artists and observers of art are just following their compression progress drive!

How does the theory explain the nature of **inductive sciences such as physics**? If the history of the entire universe were computable, and there is no evidence against this possibility (Sch06c), then its simplest explanation would be the shortest program that computes it. Unfortunately there is no general way of finding the shortest program computing any given data (LV97). Therefore physicists have traditionally proceeded incrementally, analyzing just a small aspect of the world at any given time, trying to find simple laws that allow for describing their limited observations better than the best previously known law, essentially trying to find a program that compresses the observed data better than the best previously known program. An unusually large compression breakthrough deserves the name *discovery*. For example, Newton's law of gravity can be formulated as a short piece of code which allows for substantially compressing many observation sequences involving falling apples and other objects. Although its predictive power is limited—for example, it does not explain quantum fluctuations of apple atoms—it still allows for greatly reducing the number of bits required to encode the data stream, by assigning short codes to events that are predictable with high probability (Huf52) under the assumption that the law holds. Einstein's general relativity theory yields additional compression progress as it compactly explains many previously unexplained deviations from Newton's predictions. Most physicists believe there is still room for further advances, and this is what is driving them to invent new experiments unveiling novel, previously unpublished patterns (Sch09c; Sch09b; Sch09a). When not occupied with optimizing *external* reward, physicists are also just following their compression progress drive!

More Formally

Let us formally consider a learning agent whose single life consists of discrete cycles or time steps $t = 1, 2, \ldots, T$. Its complete lifetime T may or may not be known in advance. In what follows, the value of any time-varying variable Q at time t ($1 \leq t \leq T$) will be denoted by $Q(t)$, the ordered sequence of values $Q(1), \ldots, Q(t)$ by $Q(\leq t)$, and the (possibly empty) sequence $Q(1), \ldots, Q(t-1)$ by $Q(< t)$. At any given t the agent receives a real-valued input $x(t)$ from the environment and executes a real-valued action $y(t)$ which may affect future inputs. At times $t < T$ its goal is to maximize future success or *utility*

$$u(t) = E_\mu \left[\sum_{\tau=t+1}^{T} r(\tau) \;\middle|\; h(\leq t) \right], \qquad (1)$$

where the reward $r(t)$ is a special real-valued input at time t, $h(t)$ the ordered triple $[x(t), y(t), r(t)]$ (hence $h(\leq t)$ is the known history up to t), and $E_\mu(\cdot \mid \cdot)$ denotes the conditional expectation operator with respect to some possibly unknown distribution μ from a set $\mathcal{M}$ of possible distributions. Here $\mathcal{M}$ reflects whatever is known about the possibly probabilistic reactions of the environment. For example, $\mathcal{M}$ may contain all computable distributions (Sol64; Sol78; LV97; Hut04). There is just one life, no need for predefined repeatable trials, no restriction to Markovian interfaces between sensors and environment, and the utility function implicitly takes into account the expected remaining lifespan $E_\mu(T \mid h(\leq t))$ and thus the possibility to extend it through appropriate actions (Sch05; Sch06b; Sch09d).

Recent work has led to the first reinforcement learning (RL) machines that are universal and optimal in various very general senses (Hut04; Sch02c; Sch09d). Such machines can in theory find out by themselves whether curiosity and creativity are useful or useless in a given environment, and learn to behave accordingly. In realistic settings, however, external rewards are extremely rare, and we cannot expect quick progress of this type, not even by optimal machines. But typically we can learn lots of useful behaviors even in absence of external rewards: unsupervised behaviors that just lead to predictable or compressible results and thus reflect the regularities in the environment, e. g., repeatable patterns in the world's reactions to certain action sequences. Here we argue that a bias towards exploring

previously unknown environmental regularities is *a priori* good in the real world as we know it, and should be inserted into practical AGIs, whose goal-directed learning will profit from this bias, in the sense that behaviors leading to external reward can often be quickly composed / derived from previously learnt, purely curiosity-driven behaviors. Here we shall not worry about the undeniable possibility that curiosity and creativity can actually be harmful and "kill the cat", that is, we assume the environment is benign enough. Based on experience with the real world it may be argued that this assumption is realistic. Our explorative bias greatly reduces the search space for goal-directed learning in environments where the acquisition of external reward has indeed a lot to do with easily learnable environmental regularities.

To establish this bias, in the spirit of our previous work since 1990 (Sch91b; Sch91a; SHS95; Sch97c; Sch02a; Sch06a; Sch07; Sch09c; Sch09b; Sch09a) we simply split the reward signal $r(t)$ into two scalar real-valued components: $r(t) = g(r_{ext}(t), r_{int}(t))$, where g maps pairs of real values to real values, e.g., $g(a, b) = a + b$. Here $r_{ext}(t)$ denotes traditional *external* reward provided by the environment, such as negative reward in response to bumping against a wall, or positive reward in response to reaching some teacher-given goal state. The formal theory of creativity, however, is especially interested in $r_{int}(t)$, the internal or *intrinsic* or *curiosity* or *creativity* or *aesthetic* reward, which is provided whenever the data compressor / internal world model of the agent improves in some measurable sense—for *purely creative* agents $r_{ext}(t) = 0$ for all valid t. The basic principle is essentially the one we published before in various variants (Sch91b; Sch91a; SHS95; Sch97c; Sch02a; Sch06a; Sch07; Sch09c; Sch09b; Sch09a):

> Generate intrinsic curiosity reward or creativity reward for the controller in response to improvements of the predictor or history compressor.

This is just a statement of the problem - we conceptually separate the goal (finding or creating data that can be explained / compressed / understood better) from the means of achieving the goal. Once the goal is formally specified in terms of an algorithm for computing curiosity rewards, let the controller's RL mechanism figure out how to translate such rewards into action sequences that allow the given compressor improvement algorithm to find and exploit previously unknown types of compressibility.

Computing Creativity Rewards

As pointed out above, predictors and compressors are closely related. Any type of partial predictability of the incoming sensory data stream can be exploited to improve the compressibility of the whole.

At any time t $(1 \leq t < T)$, given some compressor program p able to compress history $h(\leq t)$, let $C(p, h(\leq t))$ denote p's compression performance on $h(\leq t)$. An appropriate performance measure would be

$$C_l(p, h(\leq t)) = l(p), \tag{2}$$

where $l(p)$ denotes the length of p, measured in number of bits: the shorter p, the more algorithmic regularity and compressibility and predictability and lawfulness in the observations so far. The ultimate limit for $C_l(p, h(\leq t))$ would be $K^*(h(\leq t))$, a variant of the Kolmogorov complexity of $h(\leq t)$, namely, the length of the shortest program (for the given hardware) that computes an output starting with $h(\leq t)$ (Sol64; Kol65; LV97; Sch02b).

$C_l(p, h(\leq t))$ does not take into account the time $\tau(p, h(\leq t))$ spent by p on computing $h(\leq t)$. An alternative performance measure inspired by concepts of optimal universal search (Lev73; Sch02c; Sch04) is

$$C_{l\tau}(p, h(\leq t)) = l(p) + \log\ \tau(p, h(\leq t)). \tag{3}$$

Here compression by one bit is worth as much as runtime reduction by a factor of $\frac{1}{2}$. From an asymptotic optimality-oriented point of view this is one of the best ways of trading off storage and computation time (Lev73; Sch02c; Sch04).

In practice so far we have mostly used less universal adaptive compressors or predictors though (Sch91b; Sch91a; SHS95; Sch02a; Sch06a).

So far we have discussed measures of compressor performance, but not of performance *improvement,* which is the essential issue in our creativity-oriented context. To repeat the point made above: *The important thing are the improvements of the compressor, not its compression performance per se.* Our creativity reward in response to the compressor's progress (due to some application-dependent compressor improvement algorithm) between times t and $t + 1$ is

$$r_{int}(t+1) = f[C(p(t), h(\leq t+1)), C(p(t+1), h(\leq t+1))], \tag{4}$$

where f maps pairs of real values to real values. Various alternative progress measures are possible; most obvious is $f(a, b) = a - b$. This corresponds to a discrete time version of maximizing the first derivative of subjective data compressibility. *Note that both the old and the new compressor have to be tested on the same data, namely, the history so far.*

Asynchronous Framework for Maximizing Creativity Reward

Compare (Sch06a; Sch07; Sch09b). Let $p(t)$ denote the agent's current compressor program at time t, $s(t)$ its current controller, and do:

Controller: At any time t $(1 \leq t < T)$ do:

1. Let $s(t)$ use (parts of) history $h(\leq t)$ to select and execute $y(t + 1)$.
2. Observe $x(t + 1)$.
3. Check if there is non-zero creativity reward $r_{int}(t+1)$ provided by the asynchronously running compressor improvement algorithm (see below). If not, set $r_{int}(t + 1) = 0$.

4. Let the controller's reinforcement learning (RL) algorithm use $h(\leq t+1)$ including $r_{int}(t+1)$ (and possibly also the latest available compressed version of the observed data—see below) to obtain a new controller $s(t+1)$, in line with objective (1). Note that some actions may actually trigger learning algorithms that compute changes of the compressor and the controller's policy, such as in (Sch02a). That is, the computational cost of learning can be taken into account by the reward optimizer, and the decision when and what to learn can be learnt as well (Sch02a).

Compressor: Set p_{new} equal to the initial data compressor. Starting at time 1, repeat forever until interrupted by death at time T:

1. Set $p_{old} = p_{new}$; get current time step t and set $h_{old} = h(\leq t)$.
2. Evaluate p_{old} on h_{old}, to obtain compressor performance measure $C(p_{old}, h_{old})$. This may take many time steps.
3. Let some (possibly application-dependent) compressor improvement algorithm (such as a learning algorithm for an adaptive neural network predictor, possibly triggered by a controller action) use h_{old} to obtain a hopefully better compressor p_{new} (such as a neural net with the same size but improved predictive power and therefore improved compression performance (SH96)). Although this may take many time steps (and could be partially performed during "sleep"), p_{new} may not be optimal, due to limitations of the learning algorithm, e.g., local maxima.
4. Evaluate p_{new} on h_{old}, to obtain $C(p_{new}, h_{old})$. This may take many time steps.
5. Get current time step τ and generate creativity reward

$$r_{int}(\tau) = f[C(p_{old}, h_{old}), C(p_{new}, h_{old})], \qquad (5)$$

e.g., $f(a, b) = a - b$.

Obviously this asynchronuous scheme may cause long temporal delays between controller actions and corresponding creativity rewards. This may impose a heavy burden on the controller's RL algorithm whose task is to assign credit to past actions (to inform the controller about beginnings of compressor evaluation processes etc., we may augment its input by unique representations of such events). Nevertheless, there are RL algorithms for this purpose which are theoretically optimal in various senses (Sch06a; Sch07; Sch09c; Sch09b).

Continuous Time

In continuous time formulation, let $O(t)$ denote the state of subjective observer O at time t. The subjective simplicity or compressibility or regularity or beauty $B(D, O(t))$ of a sequence of observations and/or actions D is the negative number of bits required to encode D, given $O(t)$'s current limited prior knowledge and limited compression method. The observer-dependent and time-dependent subjective *Interestingness* or *Novelty* or *Surprise* or *Aesthetic Reward* or *Aesthetic Value* or *Joy* or *Fun* $I(D, O(t))$ is

$$I(D, O(t)) \sim \frac{\partial B(D, O(t))}{\partial t}, \qquad (6)$$

the *first derivative* of subjective simplicity: as O improves its compression algorithm, formerly apparently random data parts become subjectively more regular and beautiful, requiring fewer and fewer bits for their encoding.

Note that there are at least two ways of getting intrinsic reward: execute a learning algorithm that improves the compression of the already known data, or execute actions that generate more data, then learn to compress or understand the new data better.

How does our formal theory of creativity and curiosity generalize the traditional field of **active learning**, e.g., (Fed72)? To optimize a function may require expensive data evaluations. Active learning typically just asks which data point to evaluate next to maximize information gain (1 step look-ahead), assuming all data point evaluations are equally costly. Our more general framework takes formally into account: (1) Agents embedded in an environment where there may be arbitrary delays between experimental actions and corresponding information gains, e.g., (SHS95; Sch91a), (2) The highly environment-dependent costs of obtaining or creating not just individual data points but data *sequences* of *a priori* unknown size, (3) Arbitrary algorithmic or statistical dependencies in sequences of actions & sensory inputs, e.g., (Sch02a; Sch06a), (4) The computational cost of learning new skills, e.g., (Sch02a). Unlike previous approaches, our systems measure and maximize algorithmic (Sol64; Kol65; LV97; Sch02b) novelty (learnable but previously unknown compressibility or predictability) of self-generated, general, spatio-temporal patterns in the history of data and actions (Sch06a; Sch07; Sch09c; Sch09b).

Ongoing and Future Work

The systems described in the first publications on artificial curiosity and creativity (Sch91b; Sch91a; SHS95; Sch02a) already can be viewed as examples of implementations of a compression progress drive that encourages the discovery or creation of novel patterns, by computationally more or less limited artificial scientists or artists. To improve our previous implementations of the basic ingredients of the creativity framework (see introduction), and to build a continually growing, mostly unsupervised AGI, we will evaluate additional combinations of novel, advanced RL algorithms and adaptive compressors, and test them on humanoid robots such as the iCUB. That is, we will (A) study better practical adaptive compressors, in particular, recent, novel artificial recurrent neural networks (RNN) (HS97; SGG$^+$09) and other general yet practically feasible

methods for making predictions; (B) investigate under which conditions learning progress measures can be computed both accurately and efficiently, without frequent expensive compressor performance evaluations on the entire history so far; (C) study the applicability of recent improved RL techniques in the fields of artificial evolution, policy gradients, and others. In particular, recently there has been substantial progress in RL algorithms that are not quite as general as the universal ones (Hut04; Sch02c; Sch09d), but nevertheless capable of learning very general, program-like behavior. In particular, evolutionary methods (Rec71) can be used for training RNN, which are general computers. One especially effective family of methods uses cooperative coevolution to search the space of network components (*neurons* or individual *synapses*) instead of complete networks. The components are *coevolved* by combining them into networks, and selecting those for reproduction that participated in the best performing networks (GSM08). Other recent promising RL techniques for RNN are based on the concept of policy gradients (SMSM99; WS07; WSPS08b; RFS08; SOR$^+$08; WSPS08a).

Conclusion and Outlook

In the real world external rewards are rare. But unsupervised AGIs using additional intrinsic rewards as described in this paper will be motivated to learn many useful behaviors even in absence of external rewards, behaviors that lead to predictable or compressible results and thus reflect regularities in the environment, such as repeatable patterns in the world's reactions to certain action sequences. Often a bias towards exploring previously unknown environmental regularities through artificial curiosity / creativity is *a priori* desirable because goal-directed learning may greatly profit from it, as behaviors leading to external reward may often be rather easy to compose from previously learnt curiosity-driven behaviors. It may be possible to formally quantify this bias towards novel patterns in form of a mixture-based prior (Sol64; Sol78; LV97; Sch02c; Hut04), a weighted sum of probability distributions on sequences of actions and resulting inputs, and derive precise conditions for improved expected external reward intake. Intrinsic reward may be viewed as analogous to a *regularizer* in supervised learning, where the prior distribution on possible hypotheses greatly influences the most probable interpretation of the data in a Bayesian framework (Bis95) (for example, the well-known weight decay term of neural networks is a consequence of a Gaussian prior with zero mean for each weight). Following the introductory discussion, some of the AGIs based on the creativity principle will become scientists, artists, or comedians.

References

C. M. Bishop. *Neural networks for pattern recognition*. Oxford University Press, 1995.

V. V. Fedorov. *Theory of optimal experiments*. Academic Press, 1972.

F. J. Gomez, J. Schmidhuber, and R. Miikkulainen. Efficient non-linear control through neuroevolution. *Journal of Machine Learning Research JMLR*, 9:937–965, 2008.

S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

D. A. Huffman. A method for construction of minimum-redundancy codes. *Proceedings IRE*, 40:1098–1101, 1952.

M. Hutter. *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*. Springer, Berlin, 2004. (On J. Schmidhuber's SNF grant 20-61847).

A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1:1–11, 1965.

L. A. Levin. Universal sequential search problems. *Problems of Information Transmission*, 9(3):265–266, 1973.

M. Li and P. M. B. Vitányi. *An Introduction to Kolmogorov Complexity and its Applications (2nd edition)*. Springer, 1997.

I. Rechenberg. Evolutionsstrategie - Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Dissertation, 1971. Published 1973 by Fromman-Holzboog.

T. Rückstieß, M. Felder, and J. Schmidhuber. State-Dependent Exploration for policy gradient methods. In W. Daelemans et al., editor, *European Conference on Machine Learning (ECML) and Principles and Practice of Knowledge Discovery in Databases 2008, Part II, LNAI 5212*, pages 234–249, 2008.

J. Schmidhuber. Curious model-building control systems. In *Proceedings of the International Joint Conference on Neural Networks, Singapore*, volume 2, pages 1458–1463. IEEE press, 1991.

J. Schmidhuber. A possibility for implementing curiosity and boredom in model-building neural controllers. In J. A. Meyer and S. W. Wilson, editors, *Proc. of the International Conference on Simulation of Adaptive Behavior: From Animals to Animats*, pages 222–227. MIT Press/Bradford Books, 1991.

J. Schmidhuber. Femmes fractales, 1997.

J. Schmidhuber. Low-complexity art. *Leonardo, Journal of the International Society for the Arts, Sciences, and Technology*, 30(2):97–103, 1997.

J. Schmidhuber. What's interesting? Technical Report IDSIA-35-97, IDSIA, 1997. ftp://ftp.idsia.ch/pub/juergen/interest.ps.gz; extended abstract in Proc. Snowbird'98, Utah, 1998; see also (Sch02a).

J. Schmidhuber. Exploring the predictable. In A. Ghosh and S. Tsuitsui, editors, *Advances in Evolutionary Computing*, pages 579–612. Springer, 2002.

J. Schmidhuber. Hierarchies of generalized Kolmogorov complexities and nonenumerable universal measures computable in the limit. *International Journal of Foundations of Computer Science*, 13(4):587–612, 2002.

J. Schmidhuber. The Speed Prior: a new simplicity measure yielding near-optimal computable predictions. In J. Kivinen and R. H. Sloan, editors, *Proceedings of the 15th Annual Conference on Computational Learning Theory (COLT 2002)*, Lecture Notes in Artificial Intelligence, pages 216–228. Springer, Sydney, Australia, 2002.

J. Schmidhuber. Optimal ordered problem solver. *Machine Learning*, 54:211–254, 2004.

J. Schmidhuber. Completely self-referential optimal reinforcement learners. In W. Duch, J. Kacprzyk, E. Oja, and S. Zadrozny, editors, *Artificial Neural Networks: Biological Inspirations - ICANN 2005, LNCS 3697*, pages 223–233. Springer-Verlag Berlin Heidelberg, 2005. Plenary talk.

J. Schmidhuber. Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts. *Connection Science*, 18(2):173–187, 2006.

J. Schmidhuber. Gödel machines: Fully self-referential optimal universal self-improvers. In B. Goertzel and C. Pennachin, editors, *Artificial General Intelligence*, pages 199–226. Springer Verlag, 2006. Variant available as arXiv:cs.LO/0309048.

J. Schmidhuber. Randomness in physics. *Nature*, 439(3):392, 2006. Correspondence.

J. Schmidhuber. Simple algorithmic principles of discovery, subjective beauty, selective attention, curiosity & creativity. In *Proc. 10th Intl. Conf. on Discovery Science (DS 2007), LNAI 4755*, pages 26–38. Springer, 2007. Joint invited lecture for *ALT 2007 and DS 2007*, Sendai, Japan, 2007.

J. Schmidhuber. Art & science as by-products of the search for novel patterns, or data compressible in unknown yet learnable ways. In M. Botta, editor, *Multiple ways to design research. Research cases that reshape the design discipline*, Swiss Design Network - Et al. Edizioni, pages 98–112. Springer, 2009.

J. Schmidhuber. Driven by compression progress: A simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. In G. Pezzulo, M. V. Butz, O. Sigaud, and G. Baldassarre, editors, *Anticipatory Behavior in Adaptive Learning Systems. From Psychological Theories to Artificial Cognitive Systems*, volume 5499 of *LNCS*, pages 48–76. Springer, 2009.

J. Schmidhuber. Simple algorithmic theory of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. *SICE Journal of the Society of Instrument and Control Engineers*, 48(1):21–32, 2009.

J. Schmidhuber. Ultimate cognition *à la* Gödel. *Cognitive Computation*, 1(2):177–193, 2009.

J. Schmidhuber, A. Graves, F. J. Gomez, S. Fernandez, and S. Hochreiter. *How to Learn Programs with Artificial Recurrent Neural Networks*. Invited by Cambridge University Press, 2009. In preparation.

J. Schmidhuber and S. Heil. Sequential neural text compression. *IEEE Transactions on Neural Networks*, 7(1):142–146, 1996.

J. Storck, S. Hochreiter, and J. Schmidhuber. Reinforcement driven information acquisition in non-deterministic environments. In *Proceedings of the International Conference on Artificial Neural Networks, Paris*, volume 2, pages 159–164. EC2 & Cie, 1995.

R. S. Sutton, D. A. McAllester, S. P. Singh, and Y. Mansour. Policy gradient methods for reinforcement learning with function approximation. In S. A. Solla, T. K. Leen, and K.-R. Müller, editors, *Advances in Neural Information Processing Systems 12, [NIPS Conference, Denver, Colorado, USA, November 29 - December 4, 1999]*, pages 1057–1063. The MIT Press, 1999.

R. J. Solomonoff. A formal theory of inductive inference. Part I. *Information and Control*, 7:1–22, 1964.

R. J. Solomonoff. Complexity-based induction systems. *IEEE Transactions on Information Theory*, IT-24(5):422–432, 1978.

F. Sehnke, C. Osendorfer, T. Rückstieß, A. Graves, J. Peters, and J. Schmidhuber. Policy gradients with parameter-based exploration for control. In *Proceedings of the International Conference on Artificial Neural Networks ICANN*, 2008.

D. Wierstra and J. Schmidhuber. Policy gradient critics. In *Proceedings of the 18th European Conference on Machine Learning (ECML 2007)*, 2007.

D. Wierstra, T. Schaul, J. Peters, and J. Schmidhuber. Fitness expectation maximization. In *Proceedings of Parallel Problem Solving from Nature (PPSN 2008)*, 2008.

D. Wierstra, T. Schaul, J. Peters, and J. Schmidhuber. Natural evolution strategies. In *Congress of Evolutionary Computation (CEC 2008)*, 2008.